WALDIR CAMARGOS JÚNIOR E SILVA

ESCOLHA DE DESCRITORES MÍNIMOS E ESTABELECIMENTO DE COLEÇÕES NUCLEARES EM *Capsicum* spp.

Tese apresentada ao Programa de Pós-graduação em Agronomia, da Universidade Federal de Goiás, como requisito parcial à obtenção do título de Doutor em Agronomia, área de concentração: Genética e Melhoramento de Plantas.

Orientador:

Prof. Dr. João Batista Duarte

Goiânia, GO – Brasil
2008

Dados Internacionais de Catalogação-na-Fonte

S586e Silva, Waldir Camargos Júnior e.
Escolha de descritores mínimos e estabelecimento de coleções nucleares em *Capsicum* spp. / Waldir Camargos Júnior e Silva. - 2008.
139 f. : il. ; 30 cm.

Inclui bibliografia.
Orientação: João Batista Duarte.
Tese (doutorado) – Universidade Federal de Goiás, Escola de Agronomia e Engenharia de Alimentos, 2008.

1. Pimenta. 2. Germoplasma. 3. *Capsicum*. 4. Análise de correspondência. 5. Análise de agrupamento. I. Duarte, João Batista (orient.) II. Título.

CDU 633.842

Permitida a reprodução total ou parcial deste documento, desde que citada a fonte – O autor

WALDIR CAMARGOS JÚNIOR E SILVA

TÍTULO: "Escolha de descritores mínimos e estabelecimento de coleções nucleares em *Capsicum* spp".

Tese DEFENDIDA e APROVADA em 05 de março de 2008, pela Banca Examinadora Constituída pelos membros:

Dr. Jaison Pereira de Oliveira
Embrapa Arroz e Feijão

Prof. Lázaro José Chaves
EA/UFG

Prof. Alexandre Siqueira Guedes Coelho
ICB I/UFG

Dr. Paulo Eduardo de Melo
Embrapa Hortaliças

Prof. João Batista Duarte
Orientador – EA/UFG

Goiânia - Goiás
Brasil

“Aos meus olhos, pareço apenas um menino brincando à beira-mar, divertindo-me com o fato de encontrar de vez em quando um seixo mais liso ou uma concha mais bonita que o normal, enquanto o grande oceano da verdade permanece completamente por descobrir à minha frente.”

Isaac Newton

“O degrau de uma escada não serve simplesmente para que alguém permaneça em cima dele, destina-se a sustentar o pé de um homem pelo tempo suficiente para que ele coloque o outro um pouco mais alto.”

Thomas Huxley

A meu pai Waldir Camargos (*in memoriam*), e minha mãe, Margareth Camargos, que proporcionaram minha educação, escolar e moral, e me ensinaram que na vida é preciso ser honesto e perseverante,

DEDICO.

AGRADECIMENTOS

Agradeço a Deus, que é a inteligência suprema que nos permite estar neste espaço-tempo, vivendo e convivendo, ensinando e aprendendo.

À minha mãe, irmãs e toda minha família, agradeço todo amor, carinho, compreensão e respeito.

À Universidade Federal de Goiás, particularmente à Escola de Agronomia e Engenharia de Alimentos (EA-UFG), pela oportunidade de realização do curso de pós-graduação e pelo apóio na realização deste trabalho.

Ao professor Dr. João Batista Duarte, pela orientação, paciência, compreensão e pelos conhecimentos compartilhados na realização deste trabalho.

Aos professores Dr. Lázaro José Chaves, Dr. Alexandre Siqueira Guedes Coelho, Dr[a]. Patrícia Guimarães dos Santos Mello, Dr. Fabrizio D´Ayala Valva, Dr. José Carlos Seraphin e ao Dr. Lúcio José Vivaldi (UnB), pelos muitos ensinamentos.

À Embrapa Hortaliças, por cederem os dados utilizados na realização deste trabalho, em especial, aos pesquisadores Sabrina Isabel Costa de Carvalho, Cláudia Silva da Costa Ribeiro e Luciano de Bem Bianchetti, pela dedicação e ensinamentos compartilhados.

Ao CNPq, na qualidade de empregador, por ter me oferecido condições de concluir esse projeto.

Aos colegas de trabalho, Alice, Marcela, Lavínia, Enamar, Jussandra, Paulo Henrique, Flávio, Cláudio e Sérgio, que me apoiaram e colaboraram para que este trabalho fosse realizado.

Aos amigos e colegas do curso de pós-graduação na EA-UFG, Angel, Robertha, Tereza, Cristiane, Masuêmia, Cláudio e Wellington, com os quais tive o prazer de conviver durante esse período.

Enfim, a todos os outros que, mesmo não tendo sido aqui citados, contribuíram de alguma forma para que eu alcançasse este objetivo, os meus mais sinceros agradecimentos.

SUMÁRIO

RESUMO GERAL 8

GENERAL ABSTRACT 9

1 INTRODUÇÃO GERAL 10

2 REFERENCIAL TEÓRICO 13
2.1 ORIGEM, IMPORTÂNCIA E CARACTERÍSTICAS DO GÊNERO C*apsicum* 13
2.2 CONSERVAÇÃO E USO DE RECURSOS GENÉTICOS VEGETAIS..... 15
2.2.1 Informações gerais 15
2.2.2 Importância da caracterização dos recursos genéticos para o melhoramento de plantas 17
2.2.3 Estabelecimento de coleções nucleares 19
2.2.4 Descritores mínimos 23
2.3 ANÁLISE DE DADOS EM BANCOS DE GERMOPLASMA 24
2.3.1 Natureza dos dados 25
2.3.2 Análise multivariada para coleções de germoplasma 26
2.3.2.1 Análise em Componentes Principais (ACP) 27
2.3.2.2 Dissimilaridade genética para variáveis multicategóricas..... 28
2.3.2.3 Análise de Agrupamento 29
2.3.2.4 Análise Fatorial de Correspondência (AFC) 31

3 IDENTIFICAÇÃO DE DESCRITORES MÍNIMOS PARA ESPÉCIES E PARA O GÊNERO *Capsicum* spp. 33
RESUMO 33
ABSTRACT 34
3.1 INTRODUÇÃO 34
3.2 MATERIAL E MÉTODOS 37
3.2.1 Origem dos dados 37
3.2.2 Análises estatísticas 37
3.3 RESULTADOS E DISCUSSÃO 41
3.3.1 Descritores mínimos para *Capsicum annuum* 41
3.3.2 Descritores mínimos para *Capsicum baccatum* 46
3.3.3 Descritores mínimos para *Capsicum chinense* 51
3.3.4 Descritores mínimos para *Capsicum frutescens* 56
3.3.5 Descritores mínimos para *Capsicum* spp. 60
3.3.6 Síntese dos resultados 65
3.4 CONCLUSÕES 67

4 ESTABELECIMENTO DE COLEÇÕES NUCLEARES PARA *Capsicum* spp. A PARTIR DE ANÁLISES DE DADOS QUALITATIVOS 68
RESUMO 68
ABSTRACT 69
4.1 INTRODUÇÃO 70
4.2 MATERIAL E MÉTODOS 72
4.2.1 Origem dos dados 72
4.2.2 Análises estatísticas 74
4.3 RESULTADOS E DISCUSSÃO 77
4.3.1 Coleção nuclear para *Capsicum annuum* 77
4.3.2 Coleção nuclear para *Capsicum baccatum* 90
4.3.3 Coleção nuclear para *Capsicum chinense* 95
4.3.4 Coleção nuclear para *Capsicum frutescens* 106
4.3.5 Coleção nuclear para *Capsicum* spp. 111
4.4 CONCLUSÕES 112

5 CONCLUSÕES GERAIS 114

6 REFERÊNCIAS 115

APÊNDICES 125

RESUMO GERAL

SILVA, W. C. J. **Escolha de descritores mínimos e estabelecimento de coleções nucleares em *Capsicum* spp**. 2008. 139 f. Tese (Doutorado em Agronomia: Genética e Melhoramento de Plantas)-Escola de Agronomia e Engenharia de Alimentos, Universidade Federal de Goiás, Goiânia, 2008.[1]

A diversidade genética vegetal representa uma inestimável garantia às possíveis adversidades que estejam colocando em risco a sobrevivência das espécies biológicas. O Brasil é um dos países mais ricos em diversidade biológica de plantas, possuindo cerca de 20% de toda a biodiversidade existente no planeta e cerca de 19% dos solos agricultáveis do mundo. Os recursos genéticos são estudados em etapas bem definidas, tais como, coleta ou introdução, multiplicação, preservação/conservação, avaliação/caracterização e uso. No Brasil, a Empresa Brasileira de Pesquisa Agropecuária gerencia um sistema nacional de curadoria de recursos genéticos, em que mais de 250 mil amostras de plantas, animais e microrganismos são conservadas. As plantas do gênero *Capsicum* (pimentas e pimentões) são encontradas em todo território nacional, apresentando uma infinidade de formas, com grande diversidade fenotípica e genotípica. Praticamente todos os estados brasileiros possuem cultivares próprias, porém, muito provavelmente, parte do que se conhece pode tratar-se de mera duplicata, visto ocorrerem diversas cultivares iguais com nomes diferentes. Dessa forma, a caracterização morfológica constitui uma tarefa de grande importância para esse gênero vegetal, para evitar a conservação e manutenção de acessos de formas genômicas semelhantes em bancos de germoplasma. A Embrapa Hortaliças possui uma coleção de germoplasma de pimentas e pimentões (*Capsicum* spp.) que conta, atualmente, com cerca de dois mil e quinhentos acessos. Estes pertencem principalmente às espécies *C. annuum*, *C. baccatum, C. chinense* e *C. frutescens*. A caracterização morfológica de grande parte desses acessos (893 acessos) tem sido feita ao longo dos últimos anos, para uma série de caracteres (descritores) que permitem investigar a estrutura de sua variabilidade genética e, a partir daí, propor estratégias de manejo da coleção. Considerando-se a necessidade de estudos visando conhecer, valorizar e utilizar os acessos da coleção de *Capsicum* spp. da Embrapa Hortaliças, desenvolveu-se este trabalho com o objetivo de descrever os acessos e os descritores disponíveis nessa coleção, de forma a buscar descritores mínimos e propor coleções nucleares para o banco de germoplasma. O estudo demonstrou que é possível reduzir, em cerca de 50%, o número de descritores atualmente utilizados na caracterização dos acessos, sem prejuízo significativo à representação da variabilidade genética da coleção. A análise fatorial de correspondência demonstrou ser apropriada para a seleção de descritores mínimos, bem como para a proposição de coleções nucleares a partir de bancos de germoplasma que se encontram caracterizados, predominantemente, com dados de natureza qualitativa. As variáveis relacionadas ao fruto apresentaram grande variabilidade em toda a coleção de germoplasma, demonstrando-se que esses atributos são importantes na discriminação genética entre os acessos. O estudo revelou, também, que a origem geográfica, independentemente da espécie, tem associação direta com a divergência genética entre os acessos em *Capsicum* spp.

Palavras-chave: pimenta, germoplasma, *Capsicum*, análise de correspondência, análise de agrupamento.

[1] Orientador: Prof. Dr. João Batista Duarte. EA-UFG.

GENERAL ABSTRACT

SILVA, W. C. J. **Choice of minimum descriptors and establishment of *Capsicum* spp. core collections**. 2008. 139 f. Thesis (PhD. in Agronomy: Genetic and Plant Breeding)- Escola de Agronomia e Engenharia de Alimentos, Universidade Federal de Goiás, Goiânia, 2008.[1]

The plant genetic diversity is a valuable guarantee for possible adversities that can be risking the survival of biological species. Brazil is one of the richest countries in biodiversity of plants, with about 20% of all the existing biodiversity on Earth and around 19% of the agricultural lands in the world. Genetic resources are studied in well-defined stages, such as collecting or introduction, multiplication, preservation/conservation, evaluation/characterization and use. In Brazil, the Empresa Brasileira de Pesquisa Agropecuária manages a national system of curatorship of genetic resources in which more than 250 thousand samples of plants, animals and microorganisms are preserved. The plants of the genus *Capsicum* (chilli and sweet peppers) are found throughout the national territory, presenting countless kinds, with a great phenotypic and genotypic diversity. Practically all the brazilian states have their own cultivars, but some of them, can be in fact, considered a mere duplicate because, sometimes, the same cultivar is known under different names. Thus, the morphological characterization is a task of great importance to this plant genus so, it is possible to avoid the conservation and maintenance of accessions of similar genomic patterns in germplasm banks. Embrapa Hortaliças has a collection of germplasm of sweet and chilli peppers (*Capsicum* spp.), which amounts to about 2,500 accessions. These mainly belong to the species *C. annuum*, *C. baccatum*, *C. chinense* and *C. frutescens*. The morphological characterization of most of these acessions (893) have been carried out in recent years for a series of characters (descriptors) which allow the investigation of the structure of their genetic variability and, from then the proposition of strategies to handle the collection. Considering the need of studies aiming to know, value and use the accessions of the *Capsicum* spp. Embrapa Hortaliças collection, this research was developed with the objective of describing the accessions and the descriptors available in this collection, in order to search for minimum descriptors and propose core collections for this germplasm bank. The study demonstrated that it is possible to reduce around 50% the number of descriptors currently used in the characterization of the accessions, without significant harm to the representation of the genetic variability of the collection. The factorial correspondence analysis showed to be appropriate for the selection of minimum descriptors, as well as to the proposition of core collections from germplasm banks, characterized, predominantly, by qualitative data. Variables related to the fruit showed great variability in all the germplasm collection, demonstrating that these attributes are important in the genetic discrimination among the accessions. The study also showed that the geographical origin, regardless of the species, is directly associated with the genetic divergence among the accessions in *Capsicum* spp.

Key words: chilli pepper, germplasm, *Capsicum*, correspondence analysis, cluster analysis.

[1] Adviser: Prof. Dr. João Batista Duarte, EA-UFG.

1 INTRODUÇÃO GERAL

A diversidade genética vegetal representa uma inestimável garantia às possíveis adversidades que estejam colocando em risco a sobrevivência das espécies biológicas. É importante ressaltar que grande parte das variedades vegetais correspondentes aos alimentos mais importantes que existiam no início do século XX já se perdeu, com destaque para germoplasma de arroz, trigo, milho, aveia, cevada, feijão e ervilha. Segundo McNeely et al. (1990), há uma perspectiva de que nas próximas décadas, um número considerável de vegetais, animais e de microrganismos, muitos deles ainda desconhecidos pelo homem, serão extintos, isto é, perdidos para sempre. Assim, cada vez mais, depender-se-á de cruzamentos genéticos diversificados para se obter novas variedades e, quem sabe, até da utilização de plantas geneticamente modificadas para fortalecer e readequar as variedades existentes, de modo a assegurar a alimentação de contingentes crescentes da população mundial (FAO, 1991).

O Brasil é um dos países mais ricos em diversidade biológica de plantas, possuindo cerca de 20% de toda a biodiversidade existente no planeta e aproximadamente 19% dos solos agricultáveis do mundo. Nestes termos, a diversidade genética reveste-se de grande importância, sendo, portanto, fundamental que se disponha de caminhos seguros que permitam a caracterização e conservação da diversidade biológica (MAPA, 2004).

Os recursos genéticos são estudados em etapas bem definidas, tais como, coleta ou introdução, multiplicação, preservação/conservação, avaliação/caracterização e uso (Hawkes, 1982). Em nível internacional, o *International Plant Genetic Resources Institute* (IPGRI) vem concentrando esforços para a sistematização de estudos sobre os recursos genéticos vegetais, conscientizando os povos para a sua importância, a fim de que este patrimônio seja conservado e mantido para as gerações atuais e futuras, incluindo grande quantidade de acessos e espécies (IBPGR, 1987). No Brasil, a Embrapa Recursos Genéticos e Biotecnologia gerencia um sistema nacional de curadoria de recursos genéticos, o qual está ligado a mais de duzentos bancos de germoplasma espalhados pelo país, com mais de 250 mil amostras de plantas, animais e microrganismos conservadas (Embrapa, 2006).

A variabilidade genética está presente de forma mais expressiva nos centros de origem e de diversidade do germoplasma, ou mesmo nos programas de melhoramento via linhagens preliminares ou avançadas, cultivares elites ou primitivas (*land races*). Para sua utilização é imprescindível que os acessos, normalmente preservados em bancos ativos de germoplasma (BAG), sejam bem caracterizados e avaliados. É preciso, também, que não haja lacunas quanto à efetiva documentação e informação sobre a origem dos acessos (dados de passaporte), bem como suas características intrínsecas e potencial de uso. Assim, é importante que os acessos possuam características utilitárias desejáveis, sobretudo aos melhoristas de plantas. É fundamental também que se disponham de métodos adequados para incorporar o germoplasma exótico ao programa de melhoramento genético.

As plantas do gênero *Capsicum* (pimentas e pimentões) são encontradas em todo território nacional, apresentando uma infinidade de formas, com grande diversidade fenotípica e genotípica. Praticamente todos os estados brasileiros possuem cultivares próprias, porém, possivelmente, muito do que se conhece seja mera duplicata, visto ocorrerem diversas cultivares iguais com nomes diferentes. Dessa forma, a caracterização morfológica constitui-se em uma tarefa de grande importância nesse gênero vegetal, para se evitarem a conservação e manutenção de acessos de constituição genômica semelhante nos bancos de germoplasma.

A Embrapa Hortaliças possui uma coleção de germoplasma de pimentas e pimentões (*Capsicum* spp.) que conta, atualmente, com cerca de dois mil e quinhentos acessos. Estes pertencem, principalmente, às espécies cultivadas *C. annuum*, *C. baccatum, C. chinense* e *C. frutescens*. A caracterização morfológica de grande parte desses acessos tem sido feita, ao longo dos últimos anos, para uma série de caracteres (descritores) que permitem investigar a estrutura de sua variabilidade genética. Nesse sentido, incluem-se estudos acerca da identidade de cada acesso, pretendendo-se, por exemplo, identificar possíveis duplicatas na coleção. É importante destacar que os caracteres avaliados, em sua maioria, envolvem descritores de natureza qualitativa, sendo que a metodologia estatística para o tratamento de dados dessa natureza, provenientes de bancos de germoplasma, não se encontra ainda bem definida na literatura. O mesmo não ocorre quando se tratam de características quantitativas (Cruz & Regazzi, 1994).

Uma limitação existente para a manutenção dos bancos de germoplasma é a escassez de recursos financeiros. Sob tais condições, há necessidade de otimizar a sua aplicação, privilegiando ou priorizando a conservação de acessos estratégicos. Essa

abordagem, no contexto de recursos genéticos, recebe a denominação de coleções nucleares (Cordeiro & Abadie, 2007). Tal otimização deve contemplar também a avaliação dos acessos, muitas vezes, representada por um trabalho exaustivo de caracterização multivariada. O uso adequado da informação resultante dessa caracterização pode reduzir substancialmente esse trabalho em sua dimensão quantitativa, sem conseqüências maiores na qualidade da caracterização, via, por exemplo, o emprego de descritores mínimos.

Considerando-se a necessidade de estudos visando conhecer, valorizar e utilizar os acessos da coleção de *Capsicum* spp. da Embrapa Hortaliças, a qual dá suporte aos seus programas de melhoramento genético, desenvolveu-se este trabalho com o objetivo de descrever os acessos e os descritores disponíveis na coleção, de forma a identificar descritores mínimos e propor coleções nucleares para a coleção completa e para suas sub-coleções, estratificadas nas quatro espécies cultivadas. Vale ressaltar que a pesquisa buscou, também, identificar procedimentos estatísticos adequados para a caracterização de acessos em bancos de germoplasma vegetal, cujos descritores sejam de natureza qualitativa.

2 REFERENCIAL TEÓRICO

2.1 ORIGEM, IMPORTÂNCIA E CARACTERÍSTICAS DO GÊNERO C*apsicum*

As espécies de pimentas e pimentões pertencem à família Solanaceae e ao gênero *Capsicum*. Este gênero possui cerca de 33 espécies, classificadas de acordo com o nível de domesticação (domesticadas, semidomesticadas e silvestres). Originárias das Américas, foram introduzidas na Europa somente depois do descobrimento do Novo Mundo (século XV) (Eshbaugh, 1980). Atualmente, são consumidas por um quarto da população mundial, principalmente como condimentos. Além disso, representam um importante nicho de mercado para a agricultura brasileira e para as indústrias alimentícia, farmacêutica e cosmética (Carvalho et al., 2003).

No Brasil são encontradas as seguintes espécies e variedades pertencentes ao gênero *Capsicum*: domesticadas − *C. annuum* var. *annuum* (ex.: pimentão, pimenta americana, jalapeño), *C. baccatum* var. *pendulum* (ex.: dedo-de-moça, cambuci), *C. chinense* (ex.: pimenta de cheiro, bode, cumari-do-Pará e murupi) e *C. frutescens* (ex.: malagueta); semidomesticadas − *C. annuum* var. *glabriusculum* (ex.: pimenta de mesa), *C. baccatum* var. *baccatum* (ex.: cumari verdadeira), *C. baccatum* var. *praetermissum* (ex.:cumari verdadeira); silvestres − *C. buforum*, *C. campylopodium*, *C. dusenii*, *C. flexuosum*, *C. mirabile*, *C. parvifolium*, *C. schottianum* e *C. villosum*. As espécies silvestres não são utilizadas pelo homem (Reifschneider, 2000).

As pimentas e pimentões são cultivados no território brasileiro, tanto em regiões de clima quente, quanto nas de clima frio. A grande variabilidade genética existente pode, primeiramente, ser observada nos frutos, que apresentam diferentes formatos, coloração, tamanho e pungência (ardume e efeito picante). A pungência é exclusiva do gênero *Capsicum* e é atribuída aos alcalóides, mais especificamente a dois capsaicinóides: a capsaicina e a diidrocapsaicina. Tais alcalóides acumulam-se na superfície da placenta (tecido localizado na parte interna do fruto) e são liberados quando o fruto sofre qualquer dano físico. Segundo Reifschneider (2000), as diferentes espécies e

variedades de pimentas podem ser discriminadas por características morfológicas visualizadas nos frutos e, principalmente, nas flores (Figura 2.1).

Figura 2.1. Variabilidade entre características de flor e fruto em *Capsicum* spp. (Fonte: Embrapa Hortaliças, 2005).

O Brasil é um importante centro de diversidade do gênero *Capsicum*, pois em seu território encontram-se representantes em todos os níveis de domesticação. As pimentas e pimentões são muito apreciados pelo povo brasileiro e o seu cultivo ocupa anualmente cerca de 13 mil hectares, com uma produção de aproximadamente 280 mil toneladas de frutos. A sustentabilidade do agronegócio das pimentas e pimentões depende de cultivares com maior produtividade e qualidade, melhor adaptadas às diferentes condições climáticas do Brasil e resistentes às principais pragas e doenças. Como em qualquer espécie cultivada, o desenvolvimento de novas cultivares de *Capsicum* spp. depende da disponibilidade de variabilidade e diversidade genética, somente possível por meio de coleta, caracterização e conservação de germoplasma relacionado.

Do ponto de vista sócio-econômico, as hortaliças pertencentes ao gênero *Capsicum* estão presentes em todas as regiões do Brasil, principalmente no Sudeste e Centro-oeste. O cultivo é realizado por pequenos, médios e grandes produtores individuais ou integrados à agroindústria. Como já comentado, o cultivo de pimentas e pimentões é uma atividade significativa para o setor agrícola brasileiro, tanto para consumo *in natura* como para processamento na forma de condimentos, conservas e molho líquido.

As pimentas contribuem como fonte geradora de renda na pequena propriedade e para fixação de pessoas no meio rural, revelando, portanto, grande importância sócio-econômica. Aliado a isso, é uma cultura geradora de empregos diretos e indiretos. As grandes agroindústrias do ramo das pimentas possuem áreas de cultivo, empregando significativo número de pessoas, principalmente na época da colheita. O mercado é bastante diversificado, desde a comercialização para o consumo *in natura* e conservas caseiras, até a exportação do produto industrializado.

Os frutos de *Capsicum* são consumidos na forma *in natura* ou processados como condimentos, conservas, corantes e, também, são utilizados na composição de remédios e até na indústria bélica (aerosol de pimenta). Dentre as espécies desse gênero, o pimentão (*Capsicum annuum* L.) destaca-se entre as dez mais importantes hortaliças cultivadas no Brasil, tanto em valor, quanto em volume comercializado. Isso porque apresenta grande diversidade de formas (Blank et al., 1995) e sabores, permitindo o seu consumo *in natura*, processado em forma de conservas, molhos ou condimentos.

2.2 CONSERVAÇÃO E USO DOS RECURSOS GENÉTICOS VEGETAIS

2.2.1 Informações gerais

A diversidade biológica ou biodiversidade é representada por todas as espécies de plantas, animais, microrganismos, em interação com os ecossistemas e os processos ecológicos dos quais essas espécies fazem parte, apresentando um profundo relacionamento ecológico e evolucionário (Falk, 1990). Os recursos genéticos são definidos como a fração da biodiversidade que tem previsão de uso atual ou potencial. Assim, em plantas, compreendem as variedades tradicionais (ainda existentes em áreas menos influenciadas pelas variedades exóticas), variedades melhoradas, linhas avançadas e espécies nativas, incluindo-se os parentes selvagens das espécies cultivadas (Borém, 1997). Os recursos genéticos são portadores de genes de grande significado para o melhoramento das respectivas espécies, embora estejam ameaçados de extinção por várias causas, dependendo da espécie. Ademais, para algumas espécies, a variabilidade genética existente nos genótipos cultivados é pequena.

O germoplasma pode ser conceituado como a estrutura física vegetal, animal ou de microrganismos, dotada de caracteres hereditários, capaz de gerar um novo indivíduo, transmitindo suas características de geração em geração. De forma mais ampla, pode ser definido como o elemento ou unidade dos recursos genéticos que comporta a variabilidade genética entre e dentro da espécie, com fins de utilização para a pesquisa em geral. A diversidade contida em um germoplasma é normalmente coletada nos centros de origem das culturas, isto é, nos locais onde se iniciaram os cultivos ou nas regiões onde se desenvolveram raças locais. Assim, deve ser protegida de eventuais perdas para garantir a sua utilização (Ribeiro, 1995).

A conservação de recursos genéticos implica na manutenção de coleções *in situ*, ou seja, nos seus locais de ocorrência, ou *ex situ*, na qual a conservação das espécies é realizada fora de seu habitat. Na conservação *ex situ* podem ser mantidos indivíduos, sementes, embriões ou outras estruturas vegetais, sob diferentes condições, dependendo do material utilizado: no campo ou em casas-de-vegetação, em câmaras secas sob baixa temperatura, em meio de cultura com baixa concentração salina (conservação *in vitro*), ou criopreservadas (Borém, 1997).

Na maioria dos casos, os bancos de germoplasma funcionam como conservação *ex situ,* em que uma amostra da variabilidade genética de determinada espécie é conservada em condições artificiais, fora do habitat natural da espécie. Em geral, o armazenamento de sementes se faz em temperatura e umidade baixas. Nessas condições, a viabilidade das sementes pode ser preservada por décadas. Todas as sementes do tipo ortodoxas, que possuem baixas taxas de metabolismo e respiração, e permanecem viáveis por um longo período de tempo sob baixa umidade e temperatura, adaptam-se bem ao armazenamento. Nestes casos a longevidade do material armazenado aumenta de forma logarítmica com a redução da temperatura e umidade. Por outro lado, as sementes recalcitrantes, com altas taxas de metabolismo e respiração, não sobrevivem sob condições secas ou de alta umidade. Sua viabilidade é muito curta e estas não são conservadas por este método, uma vez que a redução da umidade abaixo de determinados níveis lhes causa danos fisiológicos. As coleções de germoplasma, reunidas por meio de expedições de coletas ou de introduções, têm sido mantidas em instituições diversas com o objetivo de conservar a diversidade genética, multiplicar e distribuir os acessos aos usuários, bem como promover a caracterização dessas coleções que são denominadas ativas.

Os bancos de germoplasma geralmente possuem duas coleções, a coleção base e a coleção ativa (ou de trabalho), utilizada mais diretamente nos programas de melhoramento. Esta última que tem um número reduzido de acessos, com características bem detalhadas. Comumente essa coleção é também denominada de coleção núcleo ou coleção nuclear (*core collection*), a qual não é fisicamente separada da coleção base, mas representa uma coleção virtual dentro da coleção completa, incluindo ampla variabilidade genética. A função primária da coleção nuclear é constituir-se na primeira fonte de variabilidade para os programas de melhoramento, de onde sairá a maioria dos acessos distribuídos pelo banco de germoplasma. A coleção base, que contem o maior número possível de acessos, deve ser utilizada quando a coleção nuclear não apresentar acessos com as características desejadas (Borém, 1997).

2.2.2 Importância da caracterização dos recursos genéticos para o melhoramento de plantas

O armazenamento da variabilidade genética em grandes coleções de germoplasma é tarefa básica para o bom andamento de qualquer programa de melhoramento. À custa deste valioso recurso, os melhoristas contam com uma espécie de reserva gênica que pode ser acionada à medida que a variabilidade vai sendo exaurida com o progresso do melhoramento. Embora muitas características genéticas tenham sido transferidas utilizando-se procedimentos clássicos de melhoramento genético, o aproveitamento do potencial oferecido pela biodiversidade poderá aumentar, sensivelmente, com o desenvolvimento e uso de modernas tecnologias que permitam, tanto superar as barreiras biológicas que separam as espécies, como permitir o aproveitamento de caracteres genéticos potenciais (Vilela-Morales & Valois, 2000).

A caracterização e a avaliação de acessos em banco de germoplasma são usadas para estabelecer diferenças ou semelhanças entre os acessos e, também, para estimular sua utilização em programas científicos e de desenvolvimento. Quando possível, devem ser realizadas em duas etapas: a caracterização dos acessos por caracteres qualitativos e a avaliação por caracteres quantitativos ou métricos, freqüentemente relacionada concomitantemente à avaliação do seu potencial de utilização (Vilela-Morales, 1988). Portanto, essas atividades envolvem ações complementares, em laboratórios e no campo, aplicadas não somente para caracterizar, avaliar e diferenciar os acessos, mas, sobretudo, para encontrar características que estimulem a utilização do germoplasma.

Com o objetivo de aumentar a precisão, a eficiência e determinar a potencialidade do uso do germoplasma, é recomendável combinar procedimentos para obtenção de diferenças morfológicas e citogenéticas, com aqueles dirigidos a encontrar qualidades e características agronômicas desejáveis à indústria e ao mercado consumidor. Vilela-Morales et al. (1993) consideram que a utilização do germoplasma será fortemente influenciada pelo conhecimento de suas características e estrutura genética, tornando-se fundamental o estudo dos seguintes aspectos: sistemática e evolução, variação genética e *pool* gênico, biologia da reprodução e barreiras reprodutivas, mapeamento genético, caracteres morfológicos, fisiológicos, de adaptação ambiental, de resistência a doenças e pragas e de interesse agrícola ou industrial, assim como a organização de linhas de pré-melhoramento.

A caracterização morfológica de um banco de germoplasma é normalmente a forma mais acessível de quantificar a sua diversidade genética, sendo bastante utilizada (Holcomb et al., 1977; Perry & McIntosh, 1991; Singh et al., 1991; Kresovich & McFerson, 1992; Jana & Singh, 1993; Li et al., 1995; Rabbani et al., 1998; Martinello et al., 2001; Silva et al., 2006). Constitui-se, portanto, em tarefa de grande importância para qualquer cultura, tanto para evitar o plantio de formas genômicas semelhantes como para evitar o estreitamento genético da espécie. Esse tipo de caracterização consiste em fornecer uma identidade para cada acesso por meio do conhecimento de uma série de dados que permitam estudar a variabilidade genética de cada amostra (Ramos & Queiroz, 1999).

Para que a diversidade genética disponível nos bancos de germoplasma seja usada no melhoramento vegetal é necessário que os acessos sejam caracterizados e documentados de forma que o melhorista possa identificar aqueles potencialmente úteis para o seu programa. Apesar da grande quantidade de acessos existentes e conservados/preservados em todo o mundo, estes não têm sido utilizados como era esperado. Peeters & Williams (1984) citam dados mostrando que, nos Estados Unidos, apenas 2% dos acessos de um banco de germoplasma de milho, bem caracterizado, foram utilizados no período de 1975-1980. Citam também que o banco de germoplasma do Centro Internacional de Arroz (IRRI), nas Filipinas, utilizou cerca de 8% dos acessos. No Brasil, a situação não é muito diferente, pois Nass et al. (1993), estudando a utilização dos recursos genéticos em milho e soja, encontraram usos de 14% e 41%, respectivamente, porém, ainda abaixo do esperado.

As conseqüências institucionais da não utilização desses recursos são sérias, pois um grande dispêndio de recursos humanos, físicos e financeiros tem sido utilizado nos últimos vinte anos para a manutenção e avaliação dessas coleções. Além disso, é provável que o pouco uso venha desencorajar o financiamento futuro, ainda mais, que a variabilidade genética contida nas coleções encontra-se a longa distância temporal da atividade agrícola, pois, quase sempre, esse germoplasma não atende às necessidades dos agricultores e consumidores. Internacionalmente, esse problema foi tratado por Brown et al. (1989) e várias causas foram apontadas. A falta de informações sobre os acessos coletados é um fator restritivo ao uso, pois os acessos têm sido caracterizados numa escala muito limitada, exceto em alguns centros internacionais (Rao et al., 1989). Os caracteres avaliados, na maioria dos casos, são baseados em descritores que não atendem às necessidades do melhorista, pois os curadores têm maior treinamento em botânica e, por conseguinte, pouca avaliação agronômica dos acessos tem sido feita. A atividade de pré-melhoramento, ainda incipiente, praticamente só começou a partir da década de 1990, e existem poucos melhoristas, especialmente em países do terceiro mundo.

O processo de caracterização consiste na anotação de descritores botânicos facilmente visíveis ou mensuráveis e que, a princípio, expressam-se em todos os ambientes (van Sloten, 1987; Burton & Davies, 1984; Williams, 1984). Dessa forma, a caracterização fixa-se, basicamente, em aspectos morfológicos e fenológicos, observados de forma sistemática por meio do uso de listas descritivas, que conduzem à discriminação dos acessos (Valls, 1988).

2.2.3 Estabelecimento de coleções nucleares

As coleções de germoplasma foram estabelecidas para preservar a diversidade genética existente, antes que muito dessa diversidade fosse perdida para sempre, devido ao uso de cultivares modernas e à exploração agrícola tecnificada. No entanto, a ênfase dada à preservação do germoplasma de culturas importantes tem levado à formação e manutenção de coleções muito grandes. Embora a representatividade da coleção possa ser conseguida por meio de coleções de tamanho grande (Frankel & Bennett, 1970), a acessibilidade e utilidade de uma coleção é inversamente relacionada com o seu tamanho (Frankel & Soulé, 1981). Basigalup et al. (1995) comentam que muitas coleções de germoplasma tornaram-se tão grandes e difusas que desencorajaram a efetiva avaliação dos acessos e,

conseqüentemente, tem atrasado a utilização do germoplasma. O excessivo tamanho dessas coleções e a limitação de recursos financeiros foram os principais fatores que contribuíram para o estabelecimento da defasagem entre o crescimento e o uso das coleções. Essa combinação de fatores tem proporcionado baixa qualidade de informações, com o aumento do tamanho das coleções, uma vez que estas não são bem caracterizadas para atributos de importância econômica.

Frankel (1984) propôs pela primeira vez o conceito de coleção nuclear (*core collection*), com o objetivo de melhorar a utilização e a acessibilidade à coleção de germoplasma. Segundo Frankel & Brown (1984), essa coleção consistiria de um número bem menor de acessos, que representaria, com um mínimo de redundância, a diversidade genética da espécie e espécies primitivas relacionadas. Aqueles acessos não incluídos na coleção nuclear seriam retidos na coleção de reserva e poderiam ser utilizados em pesquisas futuras, na busca intensiva de alelos raros, como indicado, pelas avaliações da coleção nuclear. De acordo com Malosetti & Abadie (2001), o desenvolvimento da coleção nuclear representa uma estratégia de baixo custo para melhorar a avaliação e a utilização do germoplasma e também para auxiliar as atividades do curador.

Segundo van Hintum et al. (2000), a criação de uma coleção nuclear pode ser muito simples, podendo ser feita para qualquer coleção de germoplasma, mesmo que não haja documentação completa ou dados completamente confiáveis. É necessário apenas o conhecimento básico sobre a coleção e a espécie envolvida e algum tempo para selecionar a coleção. De acordo com Brown & Spillane (1999), um procedimento geral para a seleção de uma coleção nuclear pode ser dividido em cinco passos: 1) identificar o material (coleção) que será representado; 2) decidir o tamanho da coleção nuclear; 3) dividir o grupo de materiais usados em subgrupos (estratos) distintos; 4) decidir o número de entradas por grupo; 5) escolher as entradas de cada grupo que serão incluídos na coleção nuclear. Cada um desses passos pode ser mais ou menos complexo, dependendo da informação e do procedimento usado.

Com relação ao tamanho da coleção nuclear, Brown (1989) simulou vários cenários, contemplando diferentes números e freqüências de alelos para cada loco, distribuídos na coleção. O autor conclui que uma amostra de aproximadamente 10% (em número de indivíduos) deve conter no mínimo 70% dos alelos (ou variabilidade) da coleção de base, com 95% de probabilidade. Essa conclusão é geral e se fundamenta em um modelo teórico que adota suposições sobre alguns parâmetros de genética de

populações. Ela oferece uma referência para o tamanho mínimo básico de uma coleção nuclear, que deveria conter de 5% a 10% da coleção de germoplasma e possuir não mais que 3.000 acessos, no caso de grandes coleções.

Na literatura encontra-se, ainda, que o tamanho da coleção nuclear depende muito do grau da redundância genética entre os acessos, dos recursos disponíveis para a manutenção dos acessos da coleção nuclear e da freqüência com que se realiza a regeneração dos acessos (Yonezawa et al., 1995). Esses autores recomendaram uma proporção de 20% a 30% da coleção original. Jackson et al. (1999) argumentam que, para muitas coleções, mesmo os 10% sugerido por Brown (1989) resultariam num tamanho de coleção nuclear muito maior que a coleção base de algumas espécies e que, mesmo o máximo de 3.000 acessos sugerido, ainda inviabilizaria a avaliação, por exemplo, com marcadores moleculares ou outras estratégias sofisticadas. Esses autores citam que o tamanho deveria ser variável e dependente da necessidade. Na prática, a maioria das coleções nucleares contém entre 5% e 20% dos acessos das coleções das quais foram estabelecidas e a maior tem em torno de 2.000 acessos (van Hintum et al., 2000). De maneira geral, existe correlação negativa entre o tamanho da coleção base e a sua porcentagem amostrada para formar a coleção nuclear. Coleções base com algumas centenas de acessos normalmente apresentam uma coleção nuclear em torno de 25%, enquanto coleções da ordem de dez mil acessos apresentam uma coleção nuclear em torno de 5% ou menos (Brown & Spillane, 1999).

Uma vez decidido o tamanho da coleção nuclear, o próximo passo é a escolha da estratégia de amostragem para identificar os acessos que serão incluídos na coleção. Duas estratégias de amostragem que têm sido discutidas são a amostragem ao acaso e a amostragem estratificada (Brown, 1989; Zeuli & Qualset, 1995). A amostragem aleatória simples permite que cada acesso na coleção tenha a mesma probabilidade de ser incluído na coleção nuclear. Essa estratégia é fácil de ser executada e livre de vícios. Na estratégia de amostragem aleatória estratificada, a coleção inicial é divida em grupos ou estratos não sobrepostos, seguindo-se a amostragem aleatória simples dentro de cada grupo. Esta estratégia é favorecida quando a coleção pode ser agrupada por características ou categorias importantes como: espécies, subespécies, cultivares, regiões geográficas, número de cromossomos, grupos de maturação, resistência a doenças, tolerância a estresse e outras características importantes, sobretudo ao melhorista (Brown, 1989). Segundo Erskine & Muehlbauer (1991), se for possível dividir a coleção base, mesmo em dois

grupos, uma melhor qualidade da coleção nuclear é esperada, principalmente quanto à diversidade genética. Isso é justificado, uma vez que a diversidade genética não é distribuída ao acaso nas populações. Como resultado dos processos de evolução, de domesticação e de melhoramento, a diversidade genética tem uma estrutura que pode, geralmente, ser resumida em um modelo hierárquico. Portanto, é fundamental para o desenvolvimento de coleções nucleares, a aplicação de estratégias de amostragem que identifiquem e particionem, ao máximo possível, a variação, realizando a separação dos acessos em grupos geneticamente divergentes, num processo de estratificação.

O procedimento de amostragem aleatória estratificada é recomendado por grande parte dos autores, para aumentar a conservação dos alelos comuns (Crossa et al., 1994; Brown, 1995; Crossa et al., 1995; Frankel et al., 1995; Harch et al., 1995; Zeuli & Qualset, 1995; Abadie et al., 2000). Assim, como sugerido por Brown (1989), a estratificação da coleção base tem sido, normalmente, o primeiro passo para se constituir uma coleção nuclear, devendo ser realizada de modo a maximizar a variação entre grupos e minimizar a variação dentro de grupos. Essa etapa é fundamental para o desenvolvimento de uma boa coleção nuclear (van Hintum et al., 2000).

Estabelecida a coleção nuclear, é importante avaliar a sua adequação, ou seja, avaliar a sua representatividade e o seu grau de repetitividade. Segundo Galwey (1995), confirmar se a coleção nuclear representa o domínio da população original, isto é, realizar a validação da coleção nuclear, é um problema difícil. Para avaliar o sucesso da escolha do método e o nível de representação da coleção nuclear, é necessário usar um ou mais métodos de validação. Existem vários critérios para essa avaliação, entre os quais a comparação de médias e variâncias entre a coleção nuclear e a coleção total tem sido um procedimento muito utilizado. Para van Hintum et al. (2000), uma avaliação preliminar pode ser feita por comparações de médias, amplitudes, freqüências e variâncias para caracteres específicos, entre os diferentes grupos da coleção nuclear e a coleção base. Espera-se que os intervalos permaneçam similares nas duas coleções, enquanto as médias movam-se em direção à mediana e as variâncias aumentem na coleção nuclear, comparativamente à coleção base.

Hu et al. (2000) desenvolveram uma estatística baseada na relação entre as amplitudes de valores observados nas coleções nuclear e base, para avaliar o grau de representatividade das primeiras coleções. Essa estatística foi denominada de taxa de coincidência (CR%). Os autores usaram também, para a mesma finalidade, uma estatística

obtida pela razão entre os coeficientes de variação (CV) associados às variáveis nas das duas coleções. Vasconcelos et al. (2007) também empregaram esse índice de coincidência, denominando-o de coeficiente de similaridade (CR%). Também utilizaram o coeficiente de determinação (R^2), obtido pela razão entre a soma de quadrados dos desvios entre acessos selecionados para compor a coleção nuclear e a soma de quadrados total dos desvios na coleção base.

2.2.4 Descritores mínimos

Conforme já reportado, a caracterização do germoplasma consiste, basicamente, na anotação de caracteres botânicos de alta herdabilidade, facilmente visíveis ou mensuráveis, e que se expressam consistentemente em todos os ambientes (Burton & Davies, 1984; Williams, 1984; Sloten, 1987). Assim entendida, a caracterização enfoca, basicamente, aspectos morfológicos e fenológicos, observados de forma sistemática nos acessos, por meio do confronto com listas de características descritivas ou "descritores". De acordo com Howes (1981), o descritor consiste em um atributo ou caráter que permite a distinção entre os acessos diferentes de uma mesma cultura, sendo o estado do descritor claramente definido como o valor que o descritor pode assumir. Desde 1977, o IPGRI vem promovendo o uso de descritores e estados de descritores para caracterização, avaliação e documentação dos recursos genéticos. O que ocorre é a constatação de que certos descritores geram alguns obstáculos como o não enquadramento dos acessos nos estados previstos, podendo, ainda, alguns estados variarem entre acessos além do previsto nas listas. Há também descritores que não são úteis para a quantificação da divergência genética entre os acessos, sendo, às vezes, de difícil obtenção. Dessa forma, constata-se a ineficiência de alguns descritores para o estudo da divergência entre acessos, como foi confirmado por Pereira et al. (1992), em estudos com mandioca.

O número de listas de descritores para caracterização e avaliação preliminar de germoplasma vegetal tem aumentado rapidamente na literatura mundial sobre recursos genéticos. Deve-se notar, porém, que as listas, por si só, nada agregam ao conhecimento científico sobre as culturas, mas, apenas acrescentam informações sobre diferenças já constatadas nas espécies. O valor dessas listas só surge pela aplicação às coleções de germoplasma, o que é feito, de forma consistente, em poucas instituições no mundo.

Os estados dos descritores podem variar entre acessos, além do previsto nas listas, e, eventualmente, alguns deles mostram-se invariáveis para coleções inteiras. A variabilidade intrínseca de certos acessos também pode não se encaixar nos rígidos estados originalmente previstos para alguns descritores. Na literatura, encontram-se vários exemplos de constatações da incapacidade discriminatória de caracteres utilizados para classificar cultivares ou acessos. Onyilagha (1986), trabalhando com espécies de *Dioscorea*, partiram do emprego de quarenta caracteres morfológicos e apenas quinze serviram para discriminar os acessos. Bekele et al. (1994), de um conjunto de 68 variáveis, selecionaram uma lista de 23 descritores para a caracterização de germoplasma de cacau. Strapasson et al. (2000) descartaram 86% dos descritores utilizados na caracterização de acessos de *Paspalum* sp.

Algumas listas de descritores mundialmente disponíveis, na maioria preparadas por grupos de trabalho comissionados pelo IPGRI, exigem a tomada de um número muito grande de observações, constituindo-se, sua aplicação ao germoplasma, em uma tarefa que consome muito tempo. Frankel & Brown (1984) criticaram a extensão e o nível de detalhamento de tais listas, salientando que não trazem ganhos científicos ou práticos equivalentes à pesada carga imposta aos curadores e melhoristas. Segundo esses autores, a tomada de dados morfológicos e fenológicos deve restringir-se aos descritores requeridos para a identificação de duplicatas e para o estabelecimento de coleções nucleares.

Poucos estudos têm sido implementados para avaliar a real contribuição de cada descritor na discriminação dos acessos. Segundo Pereira (1989), embora cada caráter contribua, de alguma forma, para a variabilidade, nenhum descritor sozinho pode ser responsável pela descrição de toda a variação. Dessa forma, a redução do número de descritores, com a eliminação daqueles que menos contribuem para a variabilidade, deve facilitar as interpretações, sem causar perdas consideráveis à informação.

2.3 ANÁLISE DE DADOS EM BANCOS DE GERMOPLASMA

Na caracterização dos acessos de um banco de germoplasma, é importante, que a coleção seja analisada segundo diferentes abordagens, para que os dados obtidos possam servir aos propósitos, quer do melhoramento ou para outros estudos. Resulta, daí, a grande dificuldade relacionada com o manuseio das metodologias adequadas para analisar os

dados obtidos em tais estudos. Provavelmente, pela complexidade matemática envolvida no uso desses métodos, os quais englobam, via de regra, procedimentos da análise estatística multivariada.

2.3.1 Natureza dos dados

Nas avaliações em bancos de germoplasma, particularmente no estudo da diversidade genética e de caracterizações, podem ser utilizadas características expressas por variáveis quantitativas, qualitativas ou mesmo ordinais. A esse respeito, não é incomum a indefinição quanto aos tipos de variáveis que se está trabalhando e como analisá-las, separadamente ou em conjunto. Crossa & Franco (2004) mencionam que a primeira sugestão para calcular distâncias entre duas observações, que tenham simultaneamente variáveis quantitativas e qualitativas, foi proposta por Gower (1971) e generalizada por Wishart (1986) para uso com método de agrupamento geométrico.

Para evitar classificações equivocadas das categorias de variáveis, é oportuno detalhar suas conceituações. As variáveis quantitativas são aquelas que medem relações quantitativas e, por conseguinte, podem ser mensuradas, isto é, representadas pelo seu valor numérico. Há dois tipos de variáveis quantitativas: contínuas e descontínuas (discretas). As quantitativas contínuas são aquelas que assumem quaisquer valores entre certos limites, tais como altura, diâmetro e produtividade. As variáveis discretas, por outro lado, assumem apenas valores específicos que não são observados em uma escala contínua, por causa da existência de descontinuidades (lacunas) entre os valores. Geralmente, são expressas por números inteiros.

As variáveis qualitativas são aquelas que expressam relações mutuamente exclusivas, portanto, não mensuráveis, requerendo, assim, uma codificação. Podem ser binárias ou multicategóricas. As variáveis qualitativas binárias são aquelas que exibem apenas dois estados, como a presença ou ausência de uma dada característica. As variáveis qualitativas são multicategóricas quando existem mais de duas características mutuamente exclusivas, mas que não se apresentam numa seqüência lógica de graus de atributo. Esse é o caso, por exemplo, do padrão de cor. Neste caso, a codificação pode ser feita por meio da utilização de uma variável fictícia, pela transformação de cada categoria em um caráter presença/ausência (Sneath & Sokal, 1973; Steel & Torrie, 1980; Bussab et al., 1990).

Uma variável é denominada ordinal quando pode ser distribuída em determinado número de categorias mutuamente exclusivas, que possuem uma ordenação natural. Esse é o caso de variáveis ordenadas numa seqüência lógica de magnitudes. De acordo com Sneath & Sokal (1973), estas variáveis podem ser processadas como características qualitativas.

Por envolver diversas técnicas analíticas, como métodos de agrupamento, análises em componentes principais ou variáveis canônicas, dentre outras, os procedimentos multivariados não devem ser aplicados indiscriminadamente quando diversas categorias de variáveis são o objeto de estudo. Assim, é necessário conhecer a peculiaridade dos procedimentos para a adequada avaliação de um grupo de variáveis. Neste aspecto, quando existe repetição de dados, o uso das variáveis canônicas permite, com elevada robustez, extrair informações sobre a diversidade genética presente. Por outro lado, quando não existir repetições dos dados, os componentes principais apresentam-se como a alternativa analítica mais conveniente. É oportuno ressaltar que tanto para variáveis canônicas quanto para componentes principais, quando diferentes tipos de variáveis estão presentes, cada categoria deve ser analisada separadamente das demais, atentando-se para o fato de que as variáveis qualitativas multicategóricas podem ser transformadas simplesmente em variáveis binárias, embora esta não seja uma condição essencial para a análise. Os métodos de aglomeração também apresentam peculiaridades para os diferentes tipos de variáveis, o que é implementado via medidas de distância genética, as quais apresentam especificidade para determinado tipo de variável, bem como pela forma como os dados foram obtidos, com ou sem repetição (Amaral Júnior & Thiébaut, 1998).

2.3.2 Análise multivariada para coleções de germoplasma

O termo análise multivariada refere-se à análise conjunta de diversas características simultaneamente. Em outras palavras, a análise multivariada constitui-se em um conjunto de técnicas analíticas que permite o trato das características em conjunto, possibilitando integrar as múltiplas informações extraídas das avaliações experimentais. Portanto, pode ser considerada uma análise unificadora (Das & Gupta, 1984; Johnson & Wickern, 1988).

Um campo de aplicação das técnicas multivariadas no melhoramento de plantas refere-se aos estudos relacionados às coleções de germoplasma. A identificação de

amostras repetidas (em duplicata), a eleição de descritores essenciais para a caracterização de acessos, bem como o estabelecimento de coleções nucleares normalmente são implementados baseando-se em técnicas estatísticas multivariadas (Duarte, 1997). Algumas técnicas multivariadas são muito utilizadas em trabalhos que buscam analisar as informações contidas em um grupo de dados, tais como a análise de componentes principais, de dissimilaridade genética e de agrupamento.

2.3.2.1 Análise em Componentes Principais (ACP)

A análise em componentes principais foi descrita, originalmente, por Karl Pearson, no início do século XX, na área da geometria espacial. Mais tarde, Hotelling (1933, 1936) ampliou o emprego da técnica a outros campos, o que tornou possível a sua utilização inclusive no melhoramento genético. Segundo Chatfield & Collins (1986), a análise de componentes principais objetiva transformar um conjunto de variáveis observadas num novo grupo de variáveis não correlacionadas, denominados componentes principais, que são derivados em ordem decrescente de importância. Segundo Cruz (1990), no contexto do melhoramento de plantas, o uso da análise de componentes principais permite: examinar correlações entre caracteres; resumir o conjunto de dados em outro menor e de sentido biológico mais interessante; avaliar a importância relativa de caracteres na discriminação de genótipos, fornecendo elementos para o descarte daqueles pouco discriminantes ou redundantes; construir índices que possibilitem o agrupamento de genótipos, bem como a seleção simultânea para vários caracteres; agrupar indivíduos similares mediante exame em dispersões gráficas bi ou tridimensionais, permitindo otimizar a manipulação de acessos em bancos de germoplasma, dentre outras aplicações.

Algebricamente, a análise de componentes principais consiste numa transformação aplicada às variáveis originais que resulta num novo conjunto de variáveis ortogonais e com propriedades especiais em termos de variância (Anderson, 1958). As novas variáveis representam combinações lineares daquelas originais, obtidas seqüencialmente, de forma que cada componente principal obtido explique maior parte da variabilidade dos dados do que os seguintes.

Na literatura verifica-se uma série de trabalhos usando análise de componentes principais. Pereira (1989) avaliou 280 acessos de mandioca do Banco Ativo de Germoplasma de Mandioca (BAGM) da Embrapa, para 28 descritores botânico-

agronômicos. O autor utilizou análise de componentes principais, tanto para estimar divergência genética e classificar os acessos, como para descartar descritores redundantes. Nesse trabalho foi possível o descarte de 50% dos descritores, baseando-se em suas contribuições para os primeiros componentes principais.

Faria (1994) avaliou cinqüenta acessos de feijão mulatinho do banco de germoplasma da Embrapa Arroz e Feijão (CNPAF), visando selecionar descritores, estimar a divergência e classificar os acessos em grupos de similaridade genética. Foram aplicados quatorze descritores botânico-agronômicos. Para o descarte de descritores utilizou-se a análise de componentes principais, baseando-se nos autovalores selecionados (três) e nos valores da distância de Mahalanobis. Foi possível uma redução de 57% dos descritores. Enfim, as técnicas multivariadas tem sido empregadas com êxito em vários trabalhos de avaliação de diversidade genética, de classificação de acessos em bancos de germoplasma e no descarte de descritores redundantes (Smith, 1984; Bekele et al., 1994).

2.3.2.2 Dissimilaridade genética para variáveis multicategóricas

As medidas de dissimilaridade são de grande importância em estudos de diversidade genética em que se busca caracterizar acessos visando encontrar duplicatas e mesmo utilizá-los em programas de melhoramento. Caracteres multicategóricos são comumente avaliados no melhoramento vegetal, principalmente relacionados com particularidades morfológicas e estruturais da planta, além de se ter grande interesse em certos atributos que conferem qualidade ao produto comercializado, como a forma, a coloração, o sabor do fruto, a consistência da polpa, entre outros (Cruz & Carneiro, 2003).

Quando os valores dos caracteres multicategóricos podem ser ordenados, estabelecendo-se uma escala, eles podem ser analisados como variáveis quantitativas discretas (Sneath & Sokal, 1973). Entretanto, em muitos casos, os valores de certas variáveis multicategóricas não podem ser ordenados, causando dificuldade na avaliação da dissimilaridade entre os genótipos estudados. Apesar de geralmente, se atribuírem códigos (valores) numéricos para cada categoria, não é apropriado o uso de medidas de dissimilaridade tradicionais, como a distância euclidiana, uma vez que não é possível afirmar que um par de indivíduos com valores mais discrepantes sejam mais distantes que outro par com valores mais próximos. Para resolver esse impasse, utilizam-se índices de dissimilaridade que usam as informações das variáveis multicategóricas, considerando,

inclusive, os indivíduos (genótipos) que não apresentam quaisquer das categorias de algumas dessas variáveis. Dessa forma, o índice de similaridade deve contemplar a coincidência de valores nas várias categorias, isto é, o índice pode ser estabelecido para pares de acessos, em função da concordância e discordância de categorias para as várias características consideradas (Cruz & Carneiro, 2003).

Sudré et al. (2006), trabalhando na avaliação da divergência genética entre acessos de *Capsicum* spp. e verificando a eficiência da utilização de medidas de dissimilaridade e da discriminação de genótipos em espécies, constatou que a análise de variáveis multicategóricas demonstrou ter um grande potencial de uso na caracterização e manejo de bancos de germoplasma. Concluiu, ainda, que os treze descritores qualitativos indicados pelo IPGRI, como essenciais para caracterização morfoagronômica de *Capsicum* spp., devem ser utilizados no manejo de coleções e bancos de germoplasma. Tais descritores são: cor do caule, cor da antera, cor da corola, número de flores por axila, posição da flor, hábito de crescimento, cor do fruto imaturo, cor do fruto maduro, forma do fruto, superfície do fruto, cor da semente, número de lóculos, cor da folha cotiledonar, constrição do anel do cálice e constrição do fruto (IPGRI, 1995).

2.3.2.3 Análise de Agrupamento

As análises de agrupamentos têm por finalidade reunir as amostras em vários grupos, de tal forma que exista homogeneidade dentro dos grupos e heterogeneidade entre grupos (Cruz & Regazzi, 1994). Há vários métodos de agrupamento, que se distinguem pelo tipo de resultado a ser fornecido e pelas diferentes formas de definir a proximidade entre um indivíduo e um grupo já formado, ou entre dois grupos quaisquer. Em todos os casos não se conhece, a *priori*, o número de grupos a serem estabelecidos, sendo que, diferentes métodos proporcionam diferentes resultados.

Dentre os procedimentos de agrupamento para dados qualitativos, a técnica de agrupamento de Tocher (Rao, 1952), o método hierárquico do “vizinho mais próximo” e a dispersão gráfica têm sido utilizados pela comunidade científica, por serem eficientes na discriminação genotípica (Martinello et al., 2002; Sudré et al., 2006). O método de Tocher utiliza uma matriz de dissimilaridade, sobre a qual é identificado o par de indivíduos mais similares que 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. Isso pode ser estabelecido arbitrariamente, ou adotando-se, como geralmente tem sido feito, o valor máximo da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo.

A técnica de agrupamento de Tocher tem sido usada também na seleção de acessos para compor as chamadas coleções nucleares. Nesse caso, porém, com uma modificação, que consiste em agrupar acessos com maior dissimilaridade. Por isso, a técnica tem sido denominada método de Tocher com critério de aglomeração inverso. Por meio dessa alteração, os acessos selecionados formam um grupo único, que corresponde à coleção nuclear. Por esse método, encontra-se o par de indivíduos mais dissimilares e esses formaram o grupo inicial. Para a inclusão de um outro acesso, são calculadas as distâncias entre o grupo inicial e os demais genótipos: o acesso que apresentar maior distância desse grupo será incluído. Assim se mantém a regra, até que se obtenha o número de acessos necessários para a formação da coleção nuclear (Vasconcelos et al., 2007).

Os métodos de agrupamento podem envolver matrizes de dissimilaridade de ordem elevada e o agrupamento realizado pode provocar perda de informações do grau de dissimilaridade, principalmente dos indivíduos pertencentes a um mesmo grupo. Assim, uma alternativa aos métodos de agrupamento, para avaliação da diversidade dentro de um conjunto de acessos, é a análise de dispersão gráfica, normalmente representada num espaço bi ou tridimensional. O fato questionável nessas análises é a dificuldade do estabelecimento de grupos de similaridade de forma menos subjetiva, com base na simples inspeção visual da dispersão. Há casos em que a visualização do grau de similaridade entre os acessos não é muito clara ou é distorcida em razão da reduzida dimensão da representação gráfica. Isso porque, em muitos casos, a análise não consegue resumir o complexo de informações das variáveis originais com informações representadas apenas por dois ou três eixos da análise gráfica. Assim, pelas vantagens e desvantagens, a utilização conjugada de métodos de dispersão gráfica com os de agrupamento tem sido a alternativa mais adequada em estudos de diversidade genética (Cruz & Carneiro, 2003).

2.3.2.4 Análise Fatorial de Correspondência (AFC)

As primeiras referências à análise fatorial de correspondência ou, simplesmente, análise de correspondência (AC) foram feitas por Hirschfeld (1935) e Fisher (1940). No entanto, segundo Judez (1989), foi somente nos anos de 1970 que Benzecri e Escofier demonstraram as suas propriedades algébricas. Esta análise é uma técnica multivariada para se examinarem relações geométricas do cruzamento, ou contingenciamento, de variáveis categorizadas (variáveis qualitativas). Nessa análise, uma decomposição dos dados é obtida para se estudar a sua estrutura, sem que um modelo seja hipotetizado ou que uma distribuição de probabilidade tenha sido assumida. O objetivo principal é a representação ótima da estrutura dos dados observados numa tabela de contingência, representando, graficamente, a estrutura dessa tabela e produzindo estatísticas complementares de controle, sem a necessidade de se assumirem modelos e distribuições fundamentais. Isso é obtido por meio da representação simultânea das categorias (linhas e colunas) de uma tabela de contingência sobre um gráfico, em dimensão reduzida, na forma de pontos.

A análise de correspondência pode ser considerada como um caso particular da análise de componentes principais, por ambas buscarem a redução de dimensão. O que as diferencia, portanto, é o tipo de matriz de dados a ser usada e, conseqüentemente, a métrica empregada. Outro aspecto, discutido por von der Heizden et al. (1989), é que a análise de correspondência geralmente é introduzida sem qualquer tratamento estatístico prévio, para dados categóricos, o que prova sua utilidade e flexibilidade. É, também, um meio de criar configurações associadas a tabelas de contingências, representando suas linhas por pontos no espaço, tal que a distância Euclidiana ponderada entre estes seja igual a distância qui-quadrado (χ^2) calculada entre as linhas da tabela.

Numa tabela de contingências, o conjunto de colunas indica as modalidades de uma variável, enquanto o conjunto de linhas corresponde às modalidades da outra variável. Na maioria das vezes, o interesse destas análises está nas distribuições relativas dentro das linhas e das colunas, ou seja, nos perfis linhas e perfis colunas. Geralmente, quando se aplica AC, a tabela de contingência tem um grande número de linhas e colunas, de modo que os perfis (linhas ou colunas) encontram-se num espaço de grande dimensão, tornando-se difícil observá-los diretamente. Por esse motivo, em AC, como em outras técnicas multivariadas, a preocupação é identificar um subespaço de dimensão reduzida,

preferencialmente, não mais do que duas ou três dimensões, e que estejam o mais próximo possível de todos os pontos perfis. Então, neste subespaço não se têm mais as posições originais e sim, suas projeções. As posições dos perfis neste subespaço podem ser vistas como uma aproximação das suas verdadeiras posições no espaço de origem. Neste contexto, as distâncias entre os perfis projetados passam a ser distâncias χ^2 aproximadas.

A AFC pode ser de dois tipos: a análise de correspondência simples (ACS), que tem sua aplicação a uma tabela de contingência de dupla entrada, e a análise fatorial de correspondência múltipla (AFCM), que é uma análise de correspondência simples aplicada não mais a uma tabela de contingência, mas a uma tabela disjuntiva completa (TDC). Assim, em AFCM têm-se informações mais detalhadas das contribuições em cada categoria estudada, o que não é possível na análise de correspondência simples. Ambas análises estão descritas, em detalhe, por Escofier & Pagès (1998) e Chaib Filho (2004), e podem ser implementadas em aplicativos computacionais como o SAS, SPAD, XLSTAT, R e SPSS.

Um dos objetivos da AFCM é realizar uma tipologia dos indivíduos. Essa tipologia se apóia numa noção de semelhança, tal que dois indivíduos serão bastante próximos se possuírem um grande número de modalidades em comum. Por outro lado, em grande parte de suas aplicações, os indivíduos são muito numerosos e não são conhecidos, senão pelas suas características presentes na tabela de dados. Nesse caso, os indivíduos são estudados através de classes definidas pelas variáveis. Uma análise dos indivíduos por suas classes deve ser tal que duas classes se assemelharão tanto mais, quanto mais próximos forem os perfis das partições do conjunto de modalidades.

A análise destaca os tipos de indivíduos que têm perfis semelhantes em relação aos atributos escolhidos para descrevê-los, levando-se em conta as distâncias entre os elementos de uma tabela disjuntiva completa e as relações baricêntricas. Assim, nessa análise considera-se que: a) dois indivíduos são semelhantes se eles compartilham as mesmas categorias para as variáveis sob análise; b) as categorias estarão próximas por estarem presentes nos mesmos indivíduos ou indivíduos semelhantes; e c) por definição, as categorias de uma mesma variável se excluem. Se elas estão próximas, esta proximidade se interpreta em termos de semelhança entre os indivíduos que as possuem (Vasconcelos, 2000).

3 IDENTIFICAÇÃO DE DESCRITORES MÍNIMOS PARA GERMOPLASMA DE *Capsicum* spp.

RESUMO

Neste estudo propôs-se selecionar descritores mínimos para caracterização de acessos do banco de germoplasma (BAG) de *Capsicum* spp. da Embrapa Hortaliças, o qual é caracterizado principalmente com base em dados de natureza qualitativa. Para isso, utilizou-se a análise fatorial de correspondência múltipla, combinada com a estimação de matrizes de distâncias genéticas entre acessos, para dados multicategóricos. Dos cerca de 2.500 acessos do BAG, neste estudo foram considerados 893 (420 de *C. annuum*, 106 de *C. baccatum*, 307 de *C. chinense* e 60 de *C. frutescens*), que se encontram completamente caracterizados segundo 56 descritores morfológicos. Foram propostas listas de descritores mínimos para a coleção completa de *Capsicum* spp. e para as sub-coleções relacionadas às quatro espécies cultivadas do gênero. Cerca de trinta descritores foram selecionados, separadamente, para cada coleção, garantindo-se uma correlação mínima de 0,90 entre as respectivas matrizes de distâncias obtidas com os descritores selecionados e com todos os descritores. A análise de correspondência múltipla demonstrou ser apropriada para a seleção de descritores mínimos, a partir dos dados disponíveis. As listas de descritores mínimos propostas possibilitam redução do número de descritores inicialmente considerados, em quase 50%, seja para a coleção completa ou para as coleções individuais de espécies. As características cor do filamento (CFn), cor do fruto maduro (CFM), número de lóculos (NLl), posição do fruto (PsF), procedência (Pro), pungência (Pug), superfície do fruto (SFt), comprimento da haste (CHt), comprimento da planta (CPt) e comprimento do fruto (CFt) foram eleitas como descritores mínimos, tanto para a coleção completa como para as sub-coleções específicas. Grande parte dos descritores selecionados corresponde a características do fruto, fundamentais para o melhoramento genético de *Capsicum*, haja vista sua relação com armazenamento, processamento, comercialização e consumo dos produtos obtidos a partir de pimentas e pimentões. Logo, é possível caracterizar coleções de germoplasma, de modo a atender, simultaneamente, aos critérios de representação da variabilidade genética e aos interesses mais imediatos do melhoramento.

Palavras-chave: pimenta, pimentão, análise multivariada, dados qualitativos, análise de correspondência múltipla.

ABSTRACT

IDENTIFICATION OF MINIMUM DESCRIPTORS FOR *Capsicum* spp. GERMOPLASM

This study was proposed to select minimum descriptors to the characterization of accessions of *Capsicum* spp. from the Embrapa Hortaliças germoplasm bank (BAG), which is characterized mainly by qualitative data. For that, the factorial multiple correspondence analysis was used, combined with matrices of genetic distances among accessions for multicategoric data. From about 2,500 accessions in the BAG, 893 (420 of *C. annuum*, 106 of *C. baccatum*, 307 of *C. chinense* and 60 of *C. frutescens*) were considered in this study because they were completely characterized according to 56 morphological descriptors. Lists of minimum descriptors were proposed for the whole collection of *Capsicum* spp. and for the sub-collections related to the four cultivated species of this genus. Around thirty minimum descriptors were selected for each collections, maintaining a minimum of 0,90 for the correlation between the respective matrices of distance among accessions obtained with the selected descriptors and with all the descriptors. The use of the multiple correspondence analysis showed to be appropriate for the selection of the minimum descriptors, from the available data. The proposed lists of minimum descriptors enable a reduction in the number of descriptors initially considered, around 50%, for the whole collection or for the individual collection of the species. The variables filament colour (CFn), fruit colour at mature stage (CFM), number of locules (NLl), fruit position (PsF), origin (Pro), pungency (Pug), fruit surface (SFt), stem length (CHt), plant length (CPt) and fruit length (CFt) were selected as minimum descriptors, both to the whole collection and their specific sub-collections. Most of the minimum descriptors chosen are related to the fruit characteristics, and are very important for the *Capsicum* genetic breeding, because they are related to storage, processing, marketing and consumption of the products got from sweet and chilli peppers. Thus, it is possible to characterize germplasm collections, in order to meet both the criteria of representation of genetic variability and the immediate interests of breeding programs.

Key words: chilli pepper, sweet pepper, multivariate analysis, qualitative data, multiple correspondence analysis.

3.1 INTRODUÇÃO

A diversidade genética vegetal representa uma inestimável garantia às possíveis adversidades que estejam colocando em risco a sobrevivência das espécies biológicas. O Brasil é um dos países mais ricos em diversidade biológica de plantas, possuindo cerca de 20% de toda a biodiversidade existente no planeta e cerca de 19% dos

solos agricultáveis do mundo (Vilela-Morales & Valois, 2000). Assim, a diversidade genética reveste-se de grande importância nos planos social e econômico.

A pesquisa em recursos genéticos e melhoramento vegetal é uma das atividades de inovação mais relevantes para o país, tendo produzido resultados que contribuíram significativamente para os principais ganhos qualitativos e quantitativos alcançados pela agricultura brasileira ao longo das últimas décadas. O melhoramento de plantas no Brasil está entre os melhores do mundo, com contribuições expressivas ao longo de todo o século XX, com especial destaque para a formação de recursos humanos e para o desenvolvimento de grande diversidade de plantas adaptadas às condições tropicais (Queiroz & Lopes, 2007).

O acesso, a organização e a manutenção de um importante acervo de variabilidade exótica e os significativos esforços de coleta, caracterização, conservação e promoção do uso de espécies autóctones de interesse agroalimentar ajudaram a criar as bases de sustentação para diversos programas de melhoramento genético. Métodos e processos confiáveis de caracterização e avaliação de germoplasma são fundamentais para aumentar o grau de utilização da variabilidade disponível. A eficiência dos estudos de diferenciação dos acessos entre si antecipará o potencial de cada um, ante as características genéticas desejadas. A formação e o manejo de coleções nucleares para as espécies cultivadas poderá trazer economia na manipulação de dezenas de milhares de acessos, maximizando o uso de germoplasma pelos programas de melhoramento.

A caracterização do germoplasma consiste, basicamente, na anotação de caracteres botânicos de alta herdabilidade, facilmente visíveis ou mensuráveis e que se expressam consistentemente em todos os ambientes (Burton & Davies, 1984; Williams, 1984; Sloten, 1987). Assim entendida, a caracterização enfoca, basicamente, aspectos morfológicos e fenológicos, observados de forma sistemática nos acessos, por meio do confronto com listas de características descritivas ou "descritores". Nessas listas, diferentes estados são previstos para cada descritor, com base na experiência prévia dos pesquisadores com a cultura. Em pouquíssimas, os estados são vinculados aos alelos que os determinam, mostrando, então, sua condição de marcadores genéticos morfológicos (IBPGR, 1982).

O número de listas de descritores para caracterização e avaliação preliminar de germoplasma vegetal tem aumentado rapidamente na literatura mundial sobre recursos genéticos. Deve-se notar, porém, que as listas, por si só, nada agregam ao conhecimento

científico sobre as culturas, apenas acrescentam informações sobre diferenças já constatadas nas espécies. O valor dessas listas só surge pela aplicação às coleções de germoplasma, o que é feito, de forma consistente, em poucas instituições no mundo.

Os estados dos descritores podem variar entre acessos, além do previsto nas listas, e, eventualmente, alguns deles mostram-se invariáveis para coleções inteiras. A variabilidade intrínseca de certos acessos também pode não se encaixar nos rígidos estados originalmente previstos para alguns descritores. Em suma, as listas de descritores, quase sempre publicadas *a priori*, como resultado do trabalho de comissões, só têm valor comprovado à medida que as atividades de caracterização se aproximam das etapas finais. Por vezes, pode-se constatar que muitos descritores não apresentam valor algum para a identificação de duplicatas ou para a discriminação segura dos acessos.

Na literatura encontram-se vários exemplos de constatações da incapacidade discriminatória de caracteres utilizados para classificar cultivares ou acessos. Onyilagha (1986), trabalhando com espécies de *Dioscorea*, partiram do emprego de 40 caracteres morfológicos e apenas 15 serviram para discriminar os acessos. Bekele et al. (1994), partindo de um conjunto de 68 variáveis, selecionaram uma lista de 23 descritores para a caracterização de germoplasma de cacau. Strapasson et al. (2000) descartaram 86% dos descritores utilizados na caracterização de acessos de *Paspalum* sp.

Algumas listas de descritores mundialmente disponíveis, na maioria preparadas por grupos de trabalho comissionados pelo IPGRI, exigem a tomada de um número muito grande de observações, constituindo-se, sua aplicação ao germoplasma, em uma tarefa que consome muito tempo. Frankel & Brown (1984) criticaram a extensão e o nível de detalhamento de tais listas, salientando que não trazem ganhos científicos ou práticos equivalentes à pesada carga imposta aos curadores e melhoristas. Para esses autores, a tomada de dados morfológicos e fenológicos deve se restringir aos descritores requeridos para a identificação de duplicatas e para o estabelecimento de coleções nucleares.

O uso de técnicas multivariadas tem sido eficaz não somente na identificação dos descritores de maior interesse, mas, também, no descarte daqueles que apresentam pouca relevância (Pereira, 1989). Esses procedimentos têm sido empregados em várias oportunidades, em culturas como o milho (Cruz, 1990; Ferreira, 1993), arroz (Rao et al., 1981), feijão (Oliveira, 1989; Castineiras, 1990; Faria, 1994), *Paspalum* sp. (Strapasson et al., 2000) e *Capsicum chinensis* (Luz, 2007).

O objetivo deste trabalho foi selecionar, dentre os descritores disponíveis na caracterização da coleção de *Capsicum* spp. da Embrapa Hortaliças, apenas aqueles fundamentais à discriminação segura dos acessos, tanto para a coleção completa de *Capsicum* spp., como para as sub-coleções relacionadas às suas quatro espécies cultivadas (*C. annuum*, *C. baccatum*, *C. chinense* e *C. frutescens*). Assim, será possível reduzir o trabalho de caracterização da coleção completa, uma vez que a coleta de dados basear-se-á numa lista reduzida de descritores, possibilitando, ainda, estender a caracterização para outros caracteres de interesse mais imediato do programa de melhoramento.

3.2 MATERIAL E MÉTODOS

3.2.1 Origem dos dados

Os dados utilizados para o desenvolvimento deste trabalho foram obtidos junto à Embrapa Hortaliças, que possui uma coleção de germoplasma de pimentas e pimentões (*Capsicum* spp.), atualmente, com cerca de 2.500 acessos. Desse total, foram considerados 893 acessos, oriundos de diferentes regiões do país e do exterior, que se encontram completamente caracterizados segundo 56 descritores morfológicos atualmente definidos (Tabela 3.1). As categorias de cada descritor são apresentadas em detalhe no Apêndice A. Embora a maior parte deles seja naturalmente qualitativa, observam-se alguns descritores de natureza quantitativa (ex.: comprimento da planta, diâmetro da haste, comprimento, largura e peso do fruto, entre outros). Estes, contudo, foram também categorizados com o objetivo de reunir todos num mesmo tipo de análise. Os 893 acessos estudados estão distribuídos entre as quatro espécies cultivadas do gênero *Capsicum*, sendo 420 acessos de *C. annuum*, 106 de *C. baccatum*, 307 de *C. chinense* e 60 de *C. frutescens*.

Para a avaliação do germoplasma, realizada na Embrapa Hortaliças, em Brasília – DF, os acessos foram semeados em bandeja de isopor e, após quarenta dias, cinco mudas de cada acesso foram transplantadas em telado, no espaçamento de 1,0 m x 0,60 m. As plantas foram conduzidas de acordo com as práticas agronômicas convencionais utilizadas para a cultura (Reifschneider, 2000), tendo sido tutoradas por estacas e irrigadas por gotejamento.

3.2.2 Análises estatísticas

Para seleção dos descritores mínimos foi utilizada a análise fatorial de correspondência múltipla – AFCM (Escofier & Pagès, 1992). Analogamente à conhecida análise em componentes principais (ACP), de aplicação a dados quantitativos, esta análise permite estudar uma população de *n* indivíduos, avaliados segundo *p* variáveis qualitativas. Pode-se dizer, genericamente, que apresenta a restrição de que os dados devem ter, sempre, valores positivos e, conforme Escofier & Pagès (1992), recomendam-se que as tabelas sejam retangulares, ou seja, o número de indivíduos deve ser maior que o número de características a elas associadas, o que se verificou no presente caso.

Tabela 3.1. Lista dos descritores para germoplasma de *Capsicum* spp., com seu número de categorias (*K*), atualmente adotados pela Embrapa Hortaliças, com respectivas abreviaturas.

Descritores	*K*	Abrev.	Descritores	*K*	Abrev.
Antocianina nodal	4	AtN	Formato da haste	3	FHt
Apêndice na ponta do fruto	2	APF	Formato da ponta do fruto	4	FPF
Aroma	3	Arm	Formato do fruto	6	FFt
Brotamento abaixo da primeira bifurcação	4	BtF	Hábito de crescimento	4	Hbt
Comprimento da haste	4	CHt	Largura da planta	4	LPt
Comprimento da placenta	3	CPl	Largura do fruto	5	LFt
Comprimento da planta	5	CPt	Macho-esterilidade	2	McE
Comprimento do fruto	6	CFt	Margem do cálice	3	MgC
Comprimento do pedúnculo	4	CpP	Mistura varietal	2	MVr
Constrição do anel do cálice	2	CCl	Número de flores por axila	8	NFA
Cor da antera	6	CAt	Número de lóculos	5	NLl
Cor da corola	11	CrC	Número de sementes por fruto	3	NSF
Cor da folha	8	CrF	Ombro do fruto	5	OFt
Cor da haste	3	CrH	Persistência entre fruto/pedicelo	3	PtF
Cor da mancha da corola	6	CCr	Pescoço na base do fruto	2	PcF
Cor da semente	4	CSt	Peso do fruto	6	PFt
Cor do filamento	7	CFn	Pigmento do cálice	2	Pgl
Cor do fruto imaturo	10	CFI	Pose da flor	7	PFr
Cor do fruto maduro	14	CFM	Posição do estigma	6	PEt
Densidade da folha	3	DsF	Posição do fruto	7	PsF
Densidade de ramificação	3	DsR	Procedência	12	Pro
Diâmetro da haste	4	DHt	Pubescência da folha	3	PbF
Dias para início do florescimento	4	DFl	Pubescência da haste	3	PHt
Dias para início da frutificação	4	DFt	Pungência	4	Pug
Espécie	4	Esp	Secção transversal	3	STv
Espessura da parede do fruto	6	EFt	Segregação	7	Seg
Forma da corola	3	FmC	Superfície da semente	3	SSt
Formato da folha	3	FmF	Superfície do fruto	5	SFt

As análises estatísticas utilizadas basearam-se, principalmente, na abordagem adotada por Cury (1993) e Dias (1994), fundamentada na proposta de Jolliffe (1972, 1973), substituindo-se ACP por AFCM. Conforme preconiza essa proposta, realizada a análise

com todos os descritores (análise completa), descarta-se aquele que mais contribuiu para o último eixo fatorial da análise. Com base no princípio de que a importância ou variância dos componentes principais decresce do primeiro para o último eixo, 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 deve ser a menos importante para explicar a variabilidade total e, portanto, é passível de descarte (Pereira, 1989). Para cada descarte realizado, outra análise é executada com os descritores restantes e, assim, sucessivamente até estabelecer um ordenamento dos descritores que mais contribuíram para o último eixo em cada análise. No presente caso, foram realizadas 55 análises, o que corresponde ao número total de descritores menos um (p-1). Na implementação de cada análise utilizou-se o procedimento CORRESP do sistema computacional SAS (SAS Institute, 2002), conforme rotina apresentada no Apêndice B.

Com base nos resultados da análise (AFCM) completa, isto é, incluindo todos os 56 descritores, e das outras análises (AFCM) seqüenciais, com redução sucessiva de uma destas variáveis, quatro critérios de seleção dos descritores mínimos foram avaliados: C1 – seleção com base na ordem inversa do descritor ($O_p^{'}$), de acordo com o descarte sucessivo de variáveis pela maior contribuição para o último eixo fatorial nas análises AFCM seqüenciais ($O_p^{'}$: ordem inversa do descritor para o *p-ésimo* eixo fatorial), isto é, seleção baseada na proposta de Jolliffe (1972, 1973); C2 – seleção com base na média aritmética simples (O_s) das ordens de contribuição do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da análise completa e da ordem $O_p^{'}$ já definida, tal que $O_s = (O_1 + O_2 + O_3 + O_p^{'})/4$; C3 – seleção com base numa média ponderada das ordens (O_w) anteriormente definidas, a saber, $O_w = \{[(O_1 + O_2 + O_3)/3] + O_p^{'}\}/2$; e C4 – seleção com base na média (O_z) das ordens de contribuição do descritor nos três primeiros eixos fatoriais da AFCM completa (O_1, O_2 e O_3, respectivamente), ponderadas pelas contribuições relativas destes eixos (w_1, w_2 e w_3), isto é: $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$. Inicialmente, antes de aplicar e avaliar os quatro critérios, definiu-se um tamanho uniforme para as listas de descritores mínimos a ser constituídas segundo cada critério. Esse tamanho, que variou em torno de trinta, foi assumido como o número de variáveis necessárias para explicar 90% da variabilidade associada ao primeiro eixo fatorial na AFCM completa. Isso foi definido, tendo-se como meta uma correlação igual ou

superior a 0,90, entre as matrizes de dissimilaridade dos acessos obtidas com todos os descritores e apenas com os descritores mínimos. Esse valor de correlação correspondente a coeficientes de determinação acima de 80%, que, conforme preconizam Steel & Torrie (1980), são necessários para se atingir uma acurácia seletiva ideal (Resende & Duarte, 2007).

Para avaliar a qualidade das listas de descritores mínimos geradas, dentro de cada grupo de acessos estudado (coleção completa ou coleções de espécie), inicialmente, foi estimada a matriz de distâncias entre os acessos, considerando-se todos os descritores. O índice de dissimilaridade utilizado foi aquele para variáveis multicategóricas, proposto por Cruz & Carneiro (2003): $d_{ii'} = D/(C+D)$, em que *i* e *i'* correspondem a um certo par de acessos (*i* =1, 2, ..., *n*; *i'* =1, 2, ..., *n*), C é o número de concordâncias de categoria, e D é o número de discordâncias de categoria. Da mesma forma, estimaram-se, dentro de cada grupo e critério estabelecido, as correspondentes matrizes de distâncias entre acessos, considerando-se apenas os descritores eleitos segundo cada critério. O melhor critério foi definido como aquele que resultou na maior correlação entre a matriz obtida com todos os descritores e a respectiva matriz obtida com os seus descritores mínimos. Para obtenção dessas matrizes foi utilizado o programa computacional GENES (Cruz, 2001).

Após a escolha do melhor critério e o estabelecimento da lista definitiva de descritores mínimos, com correlação mínima de 0,90 entre as matrizes de dissimilaridades já referidas, foi realizado o teste de Mantel para avaliar a correspondência entre as matrizes que representam, respectivamente, a estrutura da variabilidade na coleção antes e após o descarte de descritores. Em síntese, o teste avalia se os elementos de uma matriz (M) se correlacionam com os de outra matriz (E). Para isso, calcula-se o valor *Z*, dado por: $Z = \sum_{i,j=1}^{n} m_{ij} e_{ij}$, em que: m_{ij} e e_{ij} são elementos das matrizes M e E a serem comparadas (no caso, a matriz de distâncias entre acessos obtida com todos os descritores e aquela obtida com os descritores mínimos, respectivamente. A significância de *Z* é verificando pela probabilidade do valor observado numa distribuição nula, obtida recalculando-se valores de *Z*, diversas vezes, após aleatorizações da ordem dos elementos de uma das matrizes. A estatística *Z* possui uma relação monotônica com o coeficiente de correlação (r) de Pearson entre as matrizes (correlação matricial), de modo que ela é, de fato, utilizada para testar a significância do valor r (Manly, 1997). Para este trabalho foram usadas dez mil aleatorizações, utilizando-se o programa computacional NTSys (Rohlf, 1989).

3.3 RESULTADOS E DISCUSSÃO

3.3.1 Descritores mínimos para *Capsicum annuum*

Na Tabela 3.2 é apresentado o ordenamento decrescente dos descritores, segundo suas contribuições para o último eixo fatorial nas análises de correspondência múltipla seqüenciais, considerando-se apenas os dados relativos aos 420 acessos de *C. annuum*. Verificou-se que a variável cor da corola (CrC) foi a que apresentou maior contribuição para o último eixo fatorial, na primeira análise, realizada com todos os descritores. Assim, de acordo com Pereira (1989), considerando que a contribuição das variáveis decresce do primeiro para o último eixo, tem-se que estes explicam uma fração muito pequena da contribuição total. Então, a variável de maior coeficiente no eixo de menor autovalor deve ser a menos importante e, portanto, passível de descarte. Por esse critério, a variável cor da corola seria, portanto, descartada. No entanto, essa variável também exibiu a quarta maior contribuição para o segundo eixo fatorial, e a maior contribuição para o terceiro eixo (Tabela 3.3). Por isso, entendeu-se que seria importante considerar também a participação do descritor nesses eixos, antes de descartar esta ou outras variáveis que mostrassem comportamento similar.

Tabela 3.2. Ordenamento decrescente dos descritores morfológicos (Var) para acessos de *Capsicum annuum*, após análises de correspondência múltipla sucessivas[1] e respectivas contribuições percentuais máximas (%) para o último eixo fatorial da análise.

Análise	Var	%	Análise	Var	%	Análise	Var	%	Análise	Var	%
1	CrC	49,85	15	FPF	13,29	29	CFM	17,88	43	NSF	31,11
2	PFt	23,78	16	LPt	12,81	30	STv	12,69	44	CpP	36,75
3	FFt	21,14	17	CPl	20,90	31	PFr	16,41	45	PHt	53,44
4	PsF	29,62	18	Pro	9,66	32	PtF	24,40	46	CCr	50,00
5	LFt	66,52	19	CFt	12,08	33	APF	22,90	–	Pgl	50,00
6	CrF	45,27	20	NLl	11,40	34	AtN	23,98	–	FHt	>0,00
7	Oft	26,30	21	CFn	13,23	35	Seg	25,49	–	FmC	>0,00
8	PbF	40,91	22	CAt	15,53	36	CHt	22,93	–	MVr	>0,00
9	EFt	52,24	23	CPt	14,30	37	Arm	24,11	–	CSt	>0,00
10	DsR	12,79	24	CrH	19,99	38	DFl	25,07	–	SSt	>0,00
11	Pug	9,11	25	NFA	17,93	39	FmF	23,70	–	CCl	0,00
12	DFt	26,20	26	SFt	15,27	40	DsF	19,94	–	Esp	0,00
13	Hbt	9,53	27	MgC	25,76	41	PEt	31,60	–	McE	0,00
14	CFI	17,37	28	BtF	24,33	42	DHt	22,01	–	PcF	0,00

[1] Análises em que se eliminava, em cada ciclo, a variável com a maior contribuição para o último eixo fatorial da análise anterior.

Tabela 3.3. Descritores (Var) que mais contribuíram para os três primeiros eixos fatoriais da análise de correspondência múltipla, e respectivas contribuições (%)[1] para a variabilidade nestes eixos, em ordem decrescente, para acessos de *Capsicum annuum* do Banco de Germoplasma da Embrapa Hortaliças.

Var	Eixo 1	Var	Eixo 2	Var	Eixo 3	Var	Eixo 1	Var	Eixo 2	Var	Eixo 3
PFt	8,34	PFt	9,33	CrC	10,59	DFt	1,00	DFt	0,92	CCr	0,65
LFt	8,25	CFt	6,38	Pgl	9,18	PtF	0,97	NLl	0,87	DHt	0,63
EFt	6,85	LFt	6,11	CrF	8,92	CPt	0,94	Seg	0,76	AtN	0,61
OFt	6,76	CrC	5,88	CFI	8,05	PbF	0,88	Pet	0,65	DsF	0,61
Pug	5,50	PFr	5,88	CrH	7,62	CrF	0,85	DsF	0,65	MVr	0,54
CPl	5,30	PsF	4,73	CFn	6,77	Pgl	0,77	PbF	0,53	PHt	0,51
FFt	4,20	EFt	4,43	CFt	5,38	Seg	0,64	BtF	0,52	NSF	0,47
NLl	4,05	Pro	4,41	CAt	3,90	CpP	0,60	STv	0,48	NLl	0,45
DsR	4,00	CrF	3,92	Hbt	3,81	FmF	0,54	CPl	0,44	NFA	0,42
FPF	3,77	CFI	3,82	Pro	2,72	Arm	0,48	MVr	0,40	PbF	0,35
BtF	3,29	Hbt	3,48	EFt	2,55	PFr	0,46	PHt	0,39	DsR	0,34
Pro	3,06	FFt	3,44	LFt	2,41	FHt	0,45	Arm	0,38	Seg	0,20
MgC	2,33	CHt	2,93	DFt	2,02	CHt	0,44	FHt	0,35	PFr	0,18
DsF	2,01	CrH	2,85	PFt	1,93	CFM	0,44	DFl	0,26	APF	0,16
STv	2,01	MgC	2,82	PsF	1,79	NFA	0,41	CpP	0,24	MgC	0,14
CFt	1,97	CFM	2,70	CHt	1,73	DHt	0,25	FmF	0,22	STv	0,08
CrC	1,95	CPt	2,64	DFl	1,63	DFl	0,16	APF	0,16	CPl	0,07
PEt	1,87	OFt	2,28	Arm	1,42	MVr	0,11	FmC	0,13	FmC	0,06
CFn	1,62	SFt	2,04	CFM	1,31	FmC	0,06	PtF	0,12	FHt	0,04
Hbt	1,56	FPF	1,94	BtF	1,30	AtN	0,04	AtN	0,05	PEt	0,03
PsF	1,50	Pgl	1,59	OFt	1,24	CCr	0,03	CCr	0,02	PtF	0,02
CAt	1,47	LPt	1,25	FPF	1,20	CSt	0,02	CSt	0,02	SSt	0,02
SFt	1,47	NSF	1,23	CPt	1,10	APF	0,02	SSt	0,01	CpP	0,00
LPt	1,44	DsR	1,19	Pug	1,08	Esp	0,00	Esp	0,00	Esp	0,00
CrH	1,33	Pug	1,18	FFt	1,07	McE	0,00	NFA	0,00	McE	0,00
NSF	1,31	DHt	1,00	LPt	0,97	CCl	0,00	McE	0,00	CCl	0,00
CFI	1,15	CAt	0,99	FmF	0,97	PcF	0,00	CCl	0,00	PcF	0,00
PHt	1,08	CFn	0,99	SFt	0,76	SSt	0,00	PcF	0,00	CSt	0,00

[1] Os três eixos explicaram, juntos, 12,25% da variabilidade total captada pela análise (eixo 1: 5,94%, eixo 2: 3,54% e eixo 3: 2,77%).

Observou-se que alguns descritores não apresentaram contribuições relativas para os eixos fatoriais (Tabelas 3.2 e 3.3), ou seja, foram invariantes para esse grupo de germoplasma (acessos de *C. annuum*), tais como: constrição do anel do cálice (CCl), macho esterilidade (McE), pescoço na base do fruto (PcF) e, obviamente, espécie (Esp), pois todos os acessos do grupo estão igualmente identificados neste atributo. Esses descritores devem, obviamente, ser descartados na seleção de descritores mínimos que

representam a espécie. Em caso semelhante, Simpson et al. (1986) destacaram a inutilidade da variável cor de folha para a separação de acessos do amendoim cultivado, por ter-se mostrado invariável, em todos os 681 acessos analisados em seu trabalho.

As dez variáveis que mais contribuíram para o primeiro eixo fatorial da análise de correspondência (Tabela 3.3), foram: peso do fruto (PFt), largura do fruto (LFt), espessura da parede do fruto (EFt), ombro do fruto (OFt), pungência (Pug), comprimento da placenta (CPl), formato do fruto (FFt), número de lóculos (NLl), densidade de ramificação (DsR) e formato da ponta do fruto (FPF). As outras variáveis seguiram-nas, com contribuições cada vez menores, conforme ilustram as duas primeiras colunas dessa tabela. Deve-se ressaltar que, nesse tipo de análise, a contribuição relativa de cada eixo fatorial é bastante pequena (Tabela 3.3), haja vista o grande número de variáveis e de categorias por variável.

Verificou-se que as variáveis com maiores participações no primeiro eixo fatorial estão relacionadas a características do fruto. Isso é um aspecto importante a ser observado, visto que esses caracteres são muito utilizadas nos processos de seleção em melhoramento genético para essa espécie, principalmente no grupo dos pimentões. Essa maior participação de características relacionadas ao fruto pode ser devido, também, aos processos de coleta e domesticação, que priorizam tais características. Isso pôde ser observado também no segundo eixo fatorial, em que as dez variáveis com as maiores contribuições foram: peso do fruto (PFt), comprimento do fruto (CFt), largura do fruto (LFt), cor da corola (CrC), pose da flor (PFr), posição do fruto (PsF), espessura da parede do fruto (EFt), procedência (Pro), cor da folha (CrF) e cor do fruto imaturo (CFI).

Para o gênero *Capsicum*, caracteres como formato do fruto, cor do fruto imaturo e maduro são alguns dos caracteres importantes para o mercado consumidor. Os pimentões são comercializados na forma imatura (verde), entretanto, ocorre também demanda por frutos maduros de diferentes cores, como vermelhos, amarelos, laranjas, cremes e até roxos (Filgueira, 2000). Há regiões que preferem o fruto com formato quadrado, enquanto outras dão preferência ao tipo cônico. Cultivares com frutos de formato quadrado são pouco plantadas no Centro-Sul do Brasil, porém, são preferidas nas regiões Nordeste e Sul. Nas regiões Sudeste e Centro-Oeste, prevalecem as cultivares com frutos de formato alongado (Sudré, 2006).

No terceiro eixo fatorial, as dez variáveis mais importantes foram cor da corola (CrC), pigmento do cálice (Pgl), cor da folha (CrF), cor do fruto imaturo (CFI), cor da

haste (CrH), cor do filamento (CFn), comprimento do fruto (CFt), cor da antera (CAt), hábito de crescimento (Hbt) e procedência (Pro). Neste eixo, portanto, os caracteres relacionados ao fruto tiveram menor participação. A participação da variável procedência (Pro), ocupando, em média, a décima posição nos três primeiros eixos (Tabela 3.3), revela que a localização geográfica da origem dos acessos (Apêndice A) tem efeito substancial na divergência genética entre eles, em *C. annuum*.

No primeiro eixo fatorial, verificou-se que 50% dos descritores totalizaram quase 90% da variabilidade captada pelo eixo, ou seja, a contribuição acumulada neste eixo, da primeira variável até a 28ª (Tabela 3.3), isto é, de PFt a PHt, corresponde a aproximadamente 90%. Isso só é possível em razão de certo grau de redundância ou de associação entre as variáveis (Jolliffe, 1972, 1973), isto é, alguns dos descritores capturam informações de mesma natureza acerca da variabilidade dos acessos. Assim, para se garantir explicação mínima dessa magnitude, no primeiro e principal eixo fatorial, dever-se-ia selecionar, pelo menos, os 29 primeiros descritores. Arredondando-se para a dezena superior mais próxima, definiu-se como trinta o tamanho da lista de descritores mínimos para *C. annuum*, a ser escolhida segundo os quatro critérios apresentados. Obtidas estas listas, selecionou-se aquela proveniente do critério com a maior correlação entre sua matriz de distâncias entre acessos, gerada a partir dos descritores mínimos selecionados, e a matriz de distâncias obtida a partir de todos os descritores.

A Tabela 3.4 apresenta o ordenamento crescente dos descritores de acordo com os quatro critérios utilizados. Descritores com ordens menores são os que apresentam as melhores propriedades descritivas. Assim, para cada critério, consideraram-se como descritores mínimos as trinta variáveis com as menores ordens. A partir do teste de Mantel, verificou-se que para todos os critérios houve correlação significativa entre as matrizes M e E, com probabilidade (p-valor) inferior a 0,01. Para o critério C1 (menores contribuições para o último eixo fatorial, nas análises seqüenciais), a correlação entre a matriz de distâncias obtida com todos os descritores (M) e a matriz obtida com os descritores mínimos selecionados (E), foi de 0,84. Essa correlação, quando se utilizou o critério C2, isto é, seleção dos descritores com as menores ordens médias obtidas do ordenamento das contribuições para os três primeiros eixos fatoriais, na AFCM com todos os descritores, e do ordenamento das contribuições para o último eixo, nas análises seqüenciais, foi 0,91.

Tabela 3.4. Ordem crescente dos descritores (Var) usados para a caracterização de acessos de *Capsicum annuum* do Banco de Germoplasma da Embrapa Hortaliças, obtida a partir da aplicação de quatro critérios (C1, C2, C3 e C4)[1] para escolha da lista de descritores mínimos.

C1		C2		C3		C4		C1		C2		C3		C4	
Var	O'_p	Var	O_s	Var	O_w	Var	Oz	Var	O'_p	Var	O_s	Var	O_w	Var	Oz
CCr	1,50	CFt	13,50	Pgl	11,32	PFt	3,94	CFt	29,00	PHt	26,00	CPt	24,60	NSF	27,17
Pgl	1,50	Pgl	14,63	NSF	16,20	LFt	4,55	Pro	30,00	LPt	26,13	FPF	24,77	STv	27,38
PHt	3,00	EFt	15,00	DsF	16,52	EFt	5,96	CPl	31,00	PEt	26,38	PFt	25,23	CHt	27,50
CpP	4,00	LFt	15,00	PHt	18,13	CrC	9,77	LPt	32,00	DHt	26,50	NLl	25,40	Pet	28,97
NSF	5,00	Pro	15,00	CHt	18,60	CFt	9,92	FPF	33,00	Arm	26,75	Pug	26,87	CFM	29,04
DHt	6,00	PFt	15,50	PEt	18,92	Pro	10,39	CFI	34,00	DFt	26,75	OFt	27,10	PFr	29,48
PEt	7,00	CrH	17,00	CFt	18,97	Oft	11,89	Hbt	35,00	DsR	27,50	CrC	27,78	PHt	32,54
DsF	8,00	CrC	17,38	CrH	20,00	FFt	12,52	DFt	36,00	STv	28,13	LPt	28,00	PbF	33,93
FmF	9,00	CFI	18,75	Pro	20,07	Hbt	14,91	Pug	37,00	DFl	28,50	PtF	28,43	Arm	34,06
DFl	10,00	Hbt	18,75	DHt	20,13	Pug	15,08	DsR	38,00	FmF	29,13	AtN	28,80	Seg	34,97
Arm	11,00	CFn	19,88	BtF	20,70	PsF	15,31	EFt	39,00	CPl	29,75	CPl	28,87	DHt	35,63
CHt	12,00	CHt	20,63	MgC	21,33	FPF	15,60	PbF	40,00	Seg	29,75	PsF	29,20	FmF	36,65
Seg	13,00	CAt	21,00	Arm	22,17	CFI	16,89	OFt	41,00	CCr	32,38	FFt	29,23	Dfl	37,80
AtN	14,00	OFt	21,00	SFt	22,40	CrH	17,30	CrF	42,00	CpP	34,13	CrF	29,50	PtF	39,32
APF	15,00	FPF	21,25	PFr	22,52	CFn	18,52	LFt	43,00	AtN	35,38	DsR	30,00	MVr	40,75
PtF	16,00	BtF	21,50	CFn	22,68	CrF	19,28	PsF	44,00	PtF	35,63	DFt	30,37	FHt	41,87
PFr	17,00	PsF	21,50	EFt	22,73	BtF	19,97	FFt	45,00	PbF	36,00	APF	30,80	CPF	41,98
STv	18,00	DsF	21,63	CCr	22,75	DsR	20,12	PFt	46,00	APF	38,25	NFA	34,03	AtN	44,27
CFM	19,00	CrF	21,75	FmF	22,77	MgC	20,36	CrC	47,00	NFA	39,25	PbF	37,13	CCR	44,77
BtF	20,00	FFt	22,25	CFM	23,00	CAt	20,67	FHt	50,00	MVr	41,75	MVr	44,93	NFA	44,82
MgC	21,00	NSF	22,25	CAt	23,15	NLL	20,69	FmC	50,00	FHt	44,50	FHt	46,10	FmC	46,48
SFt	22,00	Pug	22,75	DFl	23,27	SFt	22,73	MVr	50,00	FmC	47,25	FmC	48,20	APF	47,23
NFA	23,00	SFt	22,88	CpP	23,50	Pgl	23,01	CSt	50,00	CSt	50,88	CSt	50,47	CSt	50,79
CrH	24,00	MgC	23,00	STv	23,77	DsF	23,55	SSt	50,00	SSt	50,88	SSt	50,73	SSt	51,83
CPt	25,00	CFM	23,88	LFt	24,00	CPl	23,78	Esp	54,50	CCl	54,00	CCl	54,18	CCL	53,89
CAt	26,00	CPt	24,00	Seg	24,00	LPt	23,99	McE	54,50	Esp	54,00	Esp	54,18	Esp	53,89
CFn	27,00	PFr	25,38	Hbt	24,53	CPt	25,15	CCl	54,50	McE	54,00	McE	54,18	McE	53,89
NLl	28,00	NLl	25,50	CFI	24,60	Dft	25,38	PcF	54,50	PcF	54,00	PcF	54,18	PCF	53,89

[1] C1: Ordem inversa da contribuição do descritor para o último eixo fatorial (O'_p), obtida a partir de análises de correspondência múltipla (AFCM) seqüenciais, em que se elimina, em cada uma, a variável com a maior contribuição para este eixo; C2: Ordem média (O_s) das contribuições do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da AFCM com todos os descritores, e da ordem O'_p já definida, tal que $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: Ordem média que pondera (O_w) diferentemente as ordens anteriormente definidas, segundo $O_w = \{[(O1 + O2 + O3)/3] + O'_p\}/2$; C4: Ordem média (Oz) dos três primeiros eixos fatoriais da análise AFCM completa (O1, O2 e O3, respectivamente), ponderada pela contribuição desses eixos, tal que $Oz = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$.

Para o critério C3, essa correlação foi de 0,88, isto é, quando a lista de descritores mínimos foi estabelecida segundo uma ordem média que pondera

diferentemente os ordenamentos anteriormente referidos. Por último, para o critério C4, ou seja, pela seleção dos descritores com as menores ordens médias ponderadas dos três primeiros eixos fatoriais da AFCM completa, a correlação foi de 0,91, assim como no critério C2. Assim, para esta espécie, os critérios C2 e C4 revelaram vantagens em relação aos demais. Considerando-se, porém, os resultados gerais obtidos com o estudo da coleção completa (*Capsicum* spp.), o critério C2 mostrou-se mais eficiente. Logo, adotou-se este critério para a composição da lista de descritores mínimos.

Enfim, os descritores mínimos escolhidos para caracterizar a coleção de germoplasma de *C. annuum* são: comprimento do fruto (CFt), pigmento do cálice (Pgl), espessura da parede do fruto (EFt), largura do fruto (LFt), procedência (Pro), peso do fruto (PFt), cor da haste (CrH), cor da corola (CrC), cor do fruto imaturo (CFI), hábito de crescimento (Hbt), cor do filamento (CFn), comprimento da haste (CHt), cor da antera (CAt), ombro do fruto (OFt), formato da ponta do fruto (FPF), brotamento abaixo da primeira bifurcação (BtF), posição do fruto (PsF), densidade da folha (DsF), cor da folha (CrF), formato do fruto (FFt), número de sementes por fruto (NSF), pungência (Pug), superfície do fruto (SFt), margem do cálice (MgC), cor do fruto maduro (CFM), comprimento da planta (CPt), pose da flor (PFr), número de lóculos (NLl), pubescência da haste (PHt) e largura da planta (LPt).

Dos trinta descritores escolhidos, onze são caracteres relacionados ao fruto. Como destacado anteriormente, considerando-se tratar de características comumente avaliadas pelo mercado consumidor de pimentas e pimentões, a caracterização dos acessos por estes descritores ficará diretamente vinculada aos interesses do melhorista de *Capsicum*, permitindo-lhe a identificação de potenciais genitores para os seus programas de cruzamentos. Os caracteres cor do fruto imaturo, formato, comprimento e largura do fruto estão relacionados, principalmente, com a comercialização *in natura*, enquanto a espessura da parede e o número de sementes por fruto são caracteres importantes para a venda do produto desidratado. A cor do fruto maduro é importante para ambas as formas de comercialização (Reifschneider, 2000).

3.3.2 Descritores mínimos para *Capsicum baccatum*

Considerando-se apenas os dados dos 106 acessos de *C. baccatum*, verificou-se que alguns descritores atualmente utilizados na coleção de germoplasma estudada não tiveram importância descritiva alguma para a caracterização dessa espécie, pois se mostraram invariantes (Tabelas 3.5 e 3.6). Tais descritores foram: macho-esterilidade (McE), constrição do anel do cálice (CCl), cor da semente (CSt), superfície da semente (SSt), cor da mancha da corola (CCr) e, obviamente, espécie (Esp). Assim, devido à sua inutilidade descritiva, essas variáveis não farão parte da lista de descritores mínimos para essa espécie. Segundo Daher et al. (1997), o uso de técnicas multivariadas possui a vantagem adicional de permitir a avaliação da importância de cada caráter estudado sobre a variação total disponível entre os acessos avaliados, possibilitando o descarte daqueles menos discriminantes, em razão de já estarem correlacionados com outras variáveis ou, como ocorreu neste caso, pela sua invariância.

Tabela 3.5. Ordenamento decrescente dos descritores morfológicos (Var) para acessos de *Capsicum baccatum*, após análises de correspondência múltipla sucessivas[1] e respectivas contribuições percentuais máximas (%) para o último eixo fatorial da análise.

Análise	Var	%	Análise	Var	%	Análise	Var	%	Análise	Var	%
1	LPt	8,66	15	FFt	12,27	29	CFl	26,63	43	Seg	25,70
2	LFt	12,24	16	NSF	27,36	30	CHt	46,42	44	CrF	36,36
3	CAt	10,18	17	PtF	19,74	31	SFt	14,10	45	MVr	38,12
4	OFt	8,38	18	NFA	13,40	32	NLl	14,31	46	PHt	28,67
5	Hbt	7,55	19	DsR	10,50	33	CFM	24,87	47	FmF	40,25
6	CFn	7,98	20	Pug	11,96	34	DFl	30,05	48	APF	55,58
7	EFt	12,07	21	PEt	7,69	35	FHt	19,86	49	Pgl	50,00
8	STv	10,85	22	CrH	15,01	36	CrC	44,72	–	CPl	50,00
9	PFt	10,09	23	DHt	12,96	37	PcF	23,55	–	McE	0,00
10	CFt	13,60	24	PsF	12,71	38	MgC	19,08	–	CCl	0,00
11	DsF	11,10	25	PFr	11,93	39	BtF	30,62	–	CSt	0,00
12	DFt	13,69	26	AtN	12,70	40	FmC	19,41	–	SSt	0,00
13	PbF	23,61	27	CpP	18,72	41	CPt	23,94	–	CCr	0,00
14	Pro	10,10	28	FPF	24,89	42	Arm	41,83	–	Esp	0,00

[1] Análises em que se eliminava, em cada ciclo, a variável com a maior contribuição para o último eixo fatorial da análise anterior.

Considerando-se os descritores que mais contribuíram para o último eixo fatorial nas sucessivas análises de correspondência múltipla (Tabela 3.5), verificou-se, por exemplo, que as variáveis largura da planta (LPt), largura do fruto (LFt), cor da antera (CAt), ombro do fruto (OFt) e hábito de crescimento (Hbt) foram as que apresentaram maiores contribuições para esse eixo, nas cinco primeiras análises. Assim, poderiam ser

interpretadas como as menos importantes e, conseqüentemente, candidatas ao descarte, caso fosse adotado apenas esse critério (C1). Porém, na Tabela 3.6, observa-se, por exemplo, que a variável largura do fruto (LFt) ocupa a sexta posição de contribuição no primeiro eixo fatorial e a primeira posição no segundo eixo; a variável ombro do fruto (OFt) ocupa a sexta posição nos segundo e terceiro eixos; entre outros casos. Fatos dessa natureza motivaram a avaliação dos outros dois critérios de seleção de descritores, em que se consideram, também, a participação das variáveis nos três primeiros eixos fatoriais.

Tabela 3.6. Descritores (Var) que mais contribuíram para os três primeiros eixos fatoriais da análise de correspondência múltipla, e respectivas contribuições (%)[1] para a variabilidade nestes eixos, em ordem decrescente, para acessos de *Capsicum baccatum* do Banco de Germoplasma da Embrapa Hortaliças.

Var	Eixo 1	Var	Eixo 2	Var	Eixo 3	Var	Eixo 1	Var	Eixo 2	Var	Eixo 3
PFt	5,77	LFt	9,53	CFI	7,83	CAt	1,39	DsR	0,56	LPt	1,21
CFt	5,56	FFt	9,41	Hbt	6,91	STv	1,39	MgC	0,46	APF	1,19
PtF	5,54	EFt	9,15	CFt	6,63	Arm	1,13	CrF	0,42	STv	1,10
EFt	5,54	STv	8,59	PsF	5,09	MVr	0,44	PbF	0,37	CrC	0,99
NSF	5,48	FPF	8,00	FPF	4,32	DsF	0,37	DFt	0,36	DHt	0,91
LFt	5,01	OFt	7,69	OFt	4,23	CPt	0,35	CAt	0,35	Pug	0,78
DFt	4,57	Pug	5,33	PFt	4,12	DHt	0,33	FmF	0,32	DsF	0,74
PbF	4,56	CFt	4,99	CpP	3,79	FmF	0,33	CpP	0,30	Arm	0,74
AtN	3,59	PFt	4,08	CHt	3,32	APF	0,29	AtN	0,30	NLl	0,60
NFA	3,52	Pro	3,98	CFn	3,08	OFt	0,25	DFl	0,29	LFt	0,54
PsF	3,38	PsF	3,36	SFt	3,05	BtF	0,25	DHt	0,27	AtN	0,54
FFt	3,38	NLl	3,22	DsR	2,97	CFM	0,24	Seg	0,24	PFr	0,53
Hbt	3,28	NSF	2,95	FFt	2,97	CPl	0,22	PcF	0,21	NSF	0,43
CrC	3,13	SFt	2,45	MgC	2,69	PFr	0,21	Hbt	0,19	CrF	0,25
NLl	2,91	PFr	1,31	EFt	2,66	CFI	0,19	FHt	0,18	CPl	0,25
CFn	2,78	CPl	1,18	CPt	2,63	CrF	0,18	CrH	0,15	FmF	0,17
DFl	2,72	CFM	1,10	Pro	2,46	Seg	0,16	APF	0,13	MVr	0,17
FPF	2,49	BtF	0,99	PbF	2,22	PcF	0,13	PHt	0,08	FHt	0,10
Pug	2,36	PEt	0,94	CAt	2,19	FHt	0,08	Pgl	0,02	FmC	0,05
PEt	2,10	CFI	0,88	DFl	2,13	MgC	0,04	PtF	0,01	PcF	0,02
PHt	2,08	CrC	0,88	NFA	2,00	FmC	0,02	Esp	0,00	Seg	0,00
CpP	1,99	CFn	0,87	DFt	1,87	Pgl	0,01	CCr	0,00	Esp	0,00
SFt	1,96	CPt	0,73	Pgl	1,87	Esp	0,00	McE	0,00	CrH	0,00
CrH	1,86	DsF	0,69	BtF	1,72	CCr	0,00	CCl	0,00	CCr	0,00
Pro	1,69	FmC	0,66	PHt	1,65	McE	0,00	MVr	0,00	McE	0,00
CHt	1,67	LPt	0,64	CFM	1,52	CCl	0,00	Arm	0,00	CCl	0,00
LPt	1,61	CHt	0,61	PtF	1,46	CSt	0,00	CSt	0,00	CSt	0,00
DsR	1,47	NFA	0,58	PEt	1,31	SSt	0,00	SSt	0,00	SSt	0,00

[1] Os três eixos explicaram, juntos, 19,43% da variabilidade total captada pela análise (eixo 1: 10,73%, eixo 2: 4,63% e eixo 3: 4,07%).

As dez variáveis que mais contribuíram para o primeiro eixo fatorial da análise de correspondência (Tabela 3.6) foram: peso do fruto (PFt), comprimento do fruto (CFt), persistência entre fruto/pedicelo (PtF), espessura da parede do fruto (EFt), número de sementes por fruto (NSF), largura do fruto (LFt), dias para início da frutificação (DFt), pubescência da folha (PbF), antocianina nodal (AtN) e número de flores por axila (NFA). Também para os acessos dessa espécie, as variáveis relacionadas ao fruto apresentaram grande variabilidade para esse eixo fatorial, o que realmente é característica marcante em todo o gênero *Capsicum*.

Nesse grupo de acessos, também foi observado que 50% dos descritores corresponderam a mais 90% da contribuição descritiva do primeiro eixo fatorial; isto é, a contribuição acumulada no primeiro eixo, da primeira variável (PFt) até a 28ª, densidade de ramificação (DsR), totalizou 92%. Portanto, reitera-se nesta espécie certa redundância entre os descritores, no que concerne à natureza da informação que cada um descreve acerca da variabilidade entre os acessos. Assim, adotando um critério conservador, definiu-se novamente como trinta o tamanho da lista de descritores da espécie. A partir dos ordenamentos estabelecidos segundo os quatro critérios (Tabela 3.7) e de suas respectivas listas de descritores, constatou-se que o critério que gerou a mais alta correlação (0,94) entre as matrizes de dissimilaridade dos acessos, geradas a partir de seus descritores mínimos e com todos os descritores, novamente foi o critério C2. As correlações observadas para os critérios C1, C3 e C4 foram 0,77, 0,86 e 0,64, respectivamente. Pelo teste de Mantel, verificou-se que, em todos os critérios, tais correlações foram significativas, com probabilidade (p-valor) inferior a 0,01.

Nesse sentido, adotou-se, outra vez, o critério C2 para a composição da lista definitiva de descritores mínimos para a coleção de *C. baccatum*, a saber: formato da ponta do fruto (FPF), posição do fruto (PsF), comprimento do fruto (CFt), peso do fruto (PFt), formato do fruto (FFt), espessura da parede do fruto (EFt), superfície do fruto (SFt), cor da corola (CrC), comprimento da haste (CHt), comprimento da planta (CPt), número de lóculos (NLl), cor do fruto imaturo (CFI), procedência (Pro), comprimento do pedúnculo (CpP), pungência (Pug), dias para início do florescimento (DFl), número de flores por axila (NFA), brotamento abaixo da primeira bifurcação (BtF), cor do filamento (CFn), número de sementes por fruto (NSF), largura do fruto (LFt), pubescência da folha (PbF), posição do estigma (PEt), pubescência da haste (PHt), ombro do fruto (OFt), comprimento da

placenta (CPl), cor do fruto maduro (CFM), dias para início da frutificação (DFt), densidade de ramificação (DsR) e hábito de crescimento (Hbt).

Tabela 3.7. Ordem crescente dos descritores (Var) usados para a caracterização de acessos de *Capsicum baccatum* do Banco de Germoplasma da Embrapa Hortaliças, obtida a partir da aplicação de quatro critérios (C1, C2, C3 e C4)[1] para escolha da lista de descritores mínimos.

C1		C2		C3		C4		C1		C2		C3		C4	
Var	O'_p	Var	O_s	Var	O_w	Var	Oz	Var	O'_p	Var	O_s	Var	O_w	Var	Oz
CPl	1,50	FPF	12,75	FPF	16,60	CFt	3,64	CrH	29,00	DsR	25,38	PEt	25,90	CPt	27,61
Pgl	1,50	PsF	13,38	CPl	17,03	PFt	4,16	PEt	30,00	Hbt	25,75	NFA	25,97	PHt	27,80
APF	3,00	CFt	13,50	PHt	17,70	EFt	5,79	Pug	31,00	MgC	26,25	NSF	26,13	CAt	28,37
FmF	4,00	PFt	14,75	CPt	17,77	FFt	9,45	DsR	32,00	STv	26,88	Seg	26,67	CFI	28,84
PHt	5,00	FFt	15,50	CrC	17,98	PsF	9,81	NFA	33,00	AtN	27,25	Pro	27,43	BtF	30,58
MVr	6,00	EFt	16,38	PsF	18,17	LFt	11,62	PtF	34,00	PtF	28,13	DsR	28,10	DsF	31,38
CrF	7,00	SFt	17,00	SFt	18,40	FPF	12,18	NSF	35,00	APF	28,75	PbF	28,33	CFM	31,59
Seg	8,00	CrC	20,38	NLl	19,43	NSF	14,45	FFt	36,00	FmF	29,75	PFr	29,23	CrH	34,74
Arm	9,00	CHt	20,75	BtF	19,90	PbF	15,81	Pro	37,00	Pgl	30,38	PtF	29,30	PFr	35,15
CPt	10,00	CPt	20,75	APF	20,40	CFn	16,17	PbF	38,00	PFr	30,75	DFt	29,40	CPl	35,36
FmC	11,00	NLl	20,75	FmF	20,87	Dft	16,44	DFt	39,00	CrF	31,13	PcF	29,43	DHt	35,81
BtF	12,00	CFI	21,63	DFl	20,90	NFA	16,59	DsF	40,00	Arm	32,00	CFn	30,70	MgC	36,59
MgC	13,00	Pro	22,25	CHt	21,40	Hbt	17,61	CFt	41,00	CAt	32,63	FHt	30,70	Arm	37,07
PcF	14,00	CpP	22,63	Pgl	21,83	SFt	18,34	PFt	42,00	FmC	33,00	LFt	31,00	FmF	37,27
CrC	15,00	Pug	22,75	FFt	22,30	NLL	18,89	STv	43,00	LPt	33,00	DHt	32,00	APF	37,44
FHt	16,00	DFl	23,00	CFt	22,63	PtF	19,03	EFt	44,00	DsF	33,13	STv	32,20	MVr	39,50
DFl	17,00	NFA	23,00	MgC	22,97	Pug	19,28	CFn	45,00	MVr	33,75	Hbt	32,87	CrF	40,59
CFM	18,00	BtF	23,13	CrF	23,13	CrC	19,32	Hbt	46,00	DHt	33,88	OFt	33,00	FmC	42,86
NLl	19,00	CFn	23,25	CFI	23,15	Pro	19,75	OFt	47,00	Seg	36,38	CrH	33,63	Pgl	43,52
SFt	20,00	NSF	23,50	CFM	23,30	Pet	21,44	CAt	48,00	PcF	37,25	DsF	35,33	PCF	45,23
CHt	21,00	LFt	23,63	CpP	23,55	AtN	21,73	LFt	49,00	CrH	37,38	CAt	38,10	Seg	45,38
CFI	22,00	PbF	24,00	PFt	23,63	CPF	22,52	LPt	50,00	FHt	38,00	LPt	38,60	FHt	45,84
FPF	23,00	PEt	24,25	MVr	24,08	Dfl	22,63	Esp	53,50	CCl	53,00	CCl	53,20	CCL	53,05
CpP	24,00	PHt	24,25	Arm	24,18	CHt	22,68	CCr	53,50	CCr	53,00	CCr	53,20	CCR	53,05
AtN	25,00	OFt	24,38	Pug	25,00	STv	23,74	McE	53,50	CSt	53,00	CSt	53,20	CSt	53,05
PFr	26,00	CPl	25,13	EFt	25,20	Oft	23,95	CCl	53,50	Esp	53,00	Esp	53,20	Esp	53,05
PsF	27,00	CFM	25,25	AtN	25,52	DsR	24,99	CSt	53,50	McE	53,00	McE	53,20	McE	53,05
DHt	28,00	DFt	25,38	FmC	25,73	LPt	27,18	SSt	53,50	SSt	53,00	SSt	53,20	SSt	53,05

[1] C1: Ordem inversa da contribuição do descritor para o último eixo fatorial (O'_p), obtida a partir de análises de correspondência múltipla (AFCM) seqüenciais, em que se elimina, em cada uma, a variável com a maior contribuição para este eixo; C2: Ordem média (O_s) das contribuições do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da AFCM com todos os descritores, e da ordem O'_p já definida, tal que $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: Ordem média que pondera (O_w) diferentemente as ordens anteriormente definidas, segundo $O_w = \{[(O1 + O2 + O3)/3] + O'_p\}/2$; C4: Ordem média (Oz) dos três primeiros eixos fatoriais da análise AFCM completa (O1, O2 e O3, respectivamente), ponderada pela contribuição desses eixos, tal que $Oz = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$.

A inclusão da variável procedência (Pro), mesmo não se tratando de um atributo intrínseco de caracterização dos acessos, mas um importante dado de passaporte, novamente se justifica pela sua contribuição relativa, sobretudo para o segundo eixo fatorial (Tabela 3.6). Logo, também para *C. baccatum*, a localização geográfica da origem dos acessos tem forte associação com a divergência genética expressa entre eles. De acordo com Galwey (1995), bons dados de passaporte, como, por exemplo, a origem geográfica, provavelmente não têm rival no fornecimento de informação concisa e não dispendiosa sobre a estrutura da variabilidade genética de uma coleção.

Também como observado para *C. annuum*, dos trinta descritores selecionados, doze são caracteres relacionados ao fruto, o que corrobora a importância desses atributos na discriminação genética entre acessos de *Capsicum*. Logo, a manutenção desses descritores, além de garantir parcimônia à caracterização do germoplasma, fornecerá informações de interesse imediato ao melhorista, visto que tais caracteres são diretamente vinculados às principais exigências do mercado consumidor, no que tange às cultivares e produtos comerciais derivados dessas pimentas (Sudré, 2006).

Isso mostra também que é possível caracterizar coleções de germoplasma de modo a atender, tanto aos critérios de representatividade da coleção em termos de variabilidade genética, como aos interesses mais imediatos do melhorista. Tais interesses vinculam-se principalmente à obtenção de cultivares, que são normalmente oriundos de cruzamentos cujo sucesso depende, grandemente, da escolha criteriosa dos genitores envolvidos. Nesse sentido, se a caracterização dos acessos de um banco de germoplasma levar em conta, principalmente, caracteres interessantes ao melhoramento, aumentam-se as possibilidades de utilização desse germoplasma, conforme delineam as políticas atuais sobre conservação e uso de recursos genéticos vegetais (FAO, 1991).

3.3.3 Descritores mínimos para *Capsicum chinense*

Na Tabela 3.8 é apresentado o ordenamento decrescente dos descritores, segundo suas contribuições para o último eixo fatorial, nas sucessivas análises de correspondência múltipla, considerando-se tão somente os dados dos 307 acessos de *C. chinense*. Segundo essa abordagem, no caso desta espécie, as variáveis mais importantes seriam apêndice na ponta do fruto (APF) e comprimento do pedúnculo (CpP), haja vista terem sido as últimas a serem descartadas nessas análises. Descritores como macho-

esterilidade (McE), pigmento do cálice (Pgl), cor da semente (CSt), superfície da semente (SSt) e, obviamente, espécie (Esp) não revelaram variabilidade entre os acessos, ou seja, foram invariantes, e, por isso, não apresentam utilidade descritiva para a espécie. Neste caso, a variável de menor importância, isto é, com a maior contribuição para o último eixo fatorial, na primeira análise (com todos os descritores), seria procedência (Pro). No entanto, a variável revelou também contribuição bastante significativa nos primeiros eixos fatoriais, com a segunda posição no segundo eixo e a terceira no terceiro deles (Tabela 3.9). Como já discutido, esse conflito de informações é que respaldou a avaliação de critérios alternativos de composição da lista de descritores mínimos.

Tabela 3.8. Ordenamento decrescente dos descritores morfológicos (Var) para acessos de *Capsicum chinense*, após análises de correspondência múltipla sucessivas[1] e respectivas contribuições percentuais máximas (%) para o último eixo fatorial da análise.

Análise	Var	%	Análise	Var	%	Análise	Var	%	Análise	Var	%
1	Pro	24,99	15	FFt	18,31	29	DFt	17,09	43	Arm	20,88
2	CFM	50,69	16	EFt	19,43	30	FPF	21,14	44	BtF	47,62
3	NFA	22,42	17	CrF	17,99	31	FmF	32,50	45	CPt	38,18
4	PHt	26,16	18	OFt	9,60	32	MgC	19,18	46	CFn	46,09
5	CFt	38,54	19	PsF	11,39	33	AtN	19,65	47	MVr	31,98
6	CCr	25,83	20	CrC	21,54	34	FHt	19,63	48	CCl	40,73
7	LFt	16,88	21	CrH	14,06	35	NSF	16,95	49	PcF	45,36
8	PFt	31,47	22	NLl	14,39	36	SFt	23,08	50	APF	50,00
9	DFl	25,75	23	DsR	22,33	37	DsF	22,30	–	CpP	50,00
10	STv	14,55	24	CHt	9,60	38	Seg	18,68	–	Pgl	0,00
11	CFI	13,39	25	PFr	16,74	39	DHt	30,46	–	CSt	0,00
12	Hbt	24,90	26	CPl	11,48	40	PtF	18,52	–	SSt	0,00
13	Pug	11,78	27	PEt	10,52	41	PbF	24,74	–	Esp	0,00
14	LPt	24,08	28	CAt	31,45	42	FmC	32,40	–	McE	0,00

[1] Análises em que se eliminava, em cada ciclo, a variável com a maior contribuição para o último eixo fatorial da análise anterior.

As dez variáveis que mais contribuíram para o primeiro eixo fatorial da análise de correspondência (Tabela 3.9), foram peso do fruto (PFt), largura do fruto (LFt), comprimento do fruto (EFt), secção transversal (STv), pose da flor (PFr), posição do fruto (PsF), comprimento da placenta (CPl), espessura da parede do fruto (EFt), pungência (Pug) e número de lóculos (NLl). Esse resultado reitera que variáveis relacionadas a características do fruto são importantes na discriminação genotípica também para *C. chinense,* pois apresentaram boas participações neste eixo. Essa informação também se repete no segundo e terceiro eixos fatoriais, nos quais se observam fortes contribuições das variáveis cor do fruto maduro (CFM), comprimento do fruto (CFt), formato do fruto (FFt),

peso do fruto (PFt), ombro do fruto (Oft), posição do fruto (PsF), espessura da parede do fruto (EFt) e largura do fruto (LFt).

Tabela 3.9. Descritores (Var) que mais contribuíram para os três primeiros eixos fatoriais da análise de correspondência múltipla, e respectivas contribuições (%)[1] para a variabilidade nestes eixos, em ordem decrescente, para acessos de *Capsicum chinense* do Banco de Germoplasma da Embrapa Hortaliças.

Var	Eixo 1	Var	Eixo 2	Var	Eixo 3	Var	Eixo 1	Var	Eixo 2	Var	Eixo 3
PFt	10,15	CFM	9,16	CFM	8,48	CFn	0,66	PHt	0,80	FPF	0,95
LFt	7,93	Pro	8,22	LPt	6,38	FPF	0,60	Pug	0,78	Seg	0,93
CFt	7,47	CFt	7,20	Pro	5,60	CHt	0,57	DsF	0,75	STv	0,92
STv	6,02	NLl	6,02	PFt	4,90	DHt	0,48	LFt	0,73	CrC	0,82
PFr	5,71	FFt	5,54	PFr	4,65	CCl	0,47	BtF	0,70	CFI	0,81
PsF	4,93	PFr	5,20	NLl	4,32	Seg	0,46	LPt	0,60	CFn	0,67
CPl	4,87	DFl	4,38	PsF	4,24	PtF	0,41	PcF	0,58	CrF	0,65
EFt	4,76	CpP	4,26	EFt	3,91	PEt	0,36	CPt	0,56	Pug	0,62
Pug	4,49	OFt	4,26	LFt	3,46	PcF	0,28	NSF	0,56	FmC	0,61
NLl	4,33	MgC	3,58	Hbt	3,11	CrF	0,24	STv	0,54	NFA	0,54
CFM	4,13	CAt	3,49	NSF	3,09	PHt	0,21	CrH	0,45	Arm	0,48
NSF	3,34	SFt	3,36	CFt	3,04	MgC	0,20	CPl	0,39	DsF	0,47
CAt	3,25	DHt	2,70	SFt	3,04	CrC	0,16	CrC	0,34	CPl	0,40
OFt	2,61	FHt	2,65	FFt	3,04	DsF	0,15	CCl	0,33	CrH	0,30
DFl	2,54	DFt	2,41	OFt	2,79	PbF	0,10	Arm	0,33	BtF	0,29
CpP	2,35	EFt	2,09	CAt	2,60	CPt	0,06	AtN	0,31	FmF	0,27
FFt	2,14	CFI	1,68	DsR	2,27	MVr	0,05	CFn	0,30	AtN	0,25
DFt	2,02	Hbt	1,60	PtF	2,27	BtF	0,04	DsR	0,20	MVr	0,24
Pro	1,97	PsF	1,44	PEt	2,20	FmC	0,03	PEt	0,19	CCr	0,23
CFI	1,44	FmC	1,36	CPt	2,19	Hbt	0,02	CCr	0,17	DHt	0,15
AtN	1,43	FmF	1,35	PbF	1,96	DsR	0,02	Seg	0,15	APF	0,10
SFt	1,40	NFA	1,31	FHt	1,94	CCr	0,02	MVr	0,07	PcF	0,07
FmF	1,36	FPF	1,30	MgC	1,90	APF	0,01	Esp	0,00	CCl	0,05
CrH	0,80	PtF	1,27	PHt	1,83	Esp	0,00	McE	0,00	Esp	0,00
NFA	0,80	PbF	1,18	CHt	1,69	McE	0,00	Pgl	0,00	McE	0,00
Arm	0,78	CrF	1,16	CpP	1,69	Pgl	0,00	APF	0,00	Pgl	0,00
FHt	0,72	PFt	1,15	DFl	1,46	CSt	0,00	CSt	0,00	CSt	0,00
LPt	0,66	CHt	0,85	DFt	1,13	SSt	0,00	SSt	0,00	SSt	0,00

[1] Os três eixos explicaram, juntos, 11,00% da variabilidade total captada pela análise (eixo 1: 4,57%, eixo 2: 3,32% e eixo 3: 3,11%).

A Tabela 3.10 mostra o ordenamento das variáveis, de acordo com os quatro critérios de composição da lista de descritores mínimos, utilizados neste trabalho. Escolhendo-se trinta variáveis, segundo cada critério, as correlações observadas entre as

matrizes de distâncias entre acessos, obtidas com todos os descritores e com os respectivos descritores mínimos, todas significativas ($p<0,01$), foram 0,82 para o critério C1, 0,88 para C2, 0,83 para C3 e 0,88 para C4. Logo, mais uma vez, destacaram-se os critérios C2 e C4, e, pelas mesmas razões já reportadas para *C. annuum*, optou-se pela lista de descritores mínimos obtidas a partir do critério C2.

A correspondente lista de descritores mínimos para *C. chinense* (baseada no critério C2), com correlação igual ou superior a 0,90 entre as referidas matrizes de dissimilaridade, inclui, portanto, os seguintes caracteres: pose da flor (PFr), número de lóculos (NLl), comprimento do pedúnculo (CpP), cor do fruto maduro (CFM), superfície do fruto (SFt), cor da antera (CAt), posição do fruto (PsF), comprimento do fruto (CFt), espessura da parede do fruto (EFt), ombro do fruto (OFt), formato do fruto (FFt), procedência (Pro), peso do fruto (PFt), número de sementes por fruto (NSF), formato da haste (FHt), dias para início da frutificação (DFt), largura do fruto (LFt), persistência entre fruto/pedicelo (PtF), dias para início do florescimento (DFl), margem do cálice (MgC), pubescência da folha (PbF), largura da planta (LPt), formato da ponta do fruto (FPF), diâmetro da haste (DHt), comprimento da planta (CPt), formato da folha (FmF), cor do fruto imaturo (CFI), comprimento da haste (CHt), cor do filamento (CFn), comprimento da placenta (CPl), forma da corola (FmC) e pungência (Pug).

Novamente observou-se que, dos 32 descritores selecionados, doze são caracteres relacionados ao fruto. Isso corrobora os resultados aqui obtidos para *C. annuum* e *C. baccatum*, bem como os de Teixeira (1996) e Sudré (2003), que destacam como de maior importância para a divergência genética em *Capsicum*, o comprimento, o diâmetro e o peso médio do fruto. Confirmou-se, ainda, a importância relativa da variável procedência (Pro), revelando que a origem geográfica do germoplasma tem associação direta com a divergência genética entre os acessos, também em *C. chinense*.

Tabela 3.10. Ordem crescente dos descritores (Var) usados para a caracterização de acessos de *Capsicum chinense* do Banco de Germoplasma da Embrapa Hortaliças, obtida a partir da aplicação de quatro critérios (C1, C2, C3 e C4)[1] para escolha da lista de descritores mínimos.

C1		C2		C3		C4		C1		C2		C3		C4	
Var	O'_p	Var	O_s	Var	O_w	Var	Oz	Var	O'_p	Var	O_s	Var	O_w	Var	Oz
CpP	1,50	PFr	10,75	CpP	8,77	CFM	5,15	DsR	29,00	CFn	28,38	APF	26,40	CHt	28,54
APF	1,50	NLl	12,50	SFt	16,13	PFr	5,30	NLl	30,00	CPl	28,50	DsF	26,40	Hbt	28,62
PcF	3,00	CpP	12,88	PFr	16,17	CFt	5,83	CrH	31,00	FmC	28,50	CPl	26,53	DHt	30,79
CCl	4,00	CFM	15,75	NSF	18,45	NLL	7,06	CrC	32,00	Pug	28,50	AtN	27,03	PbF	31,35
MVr	5,00	SFt	15,75	NLL	18,47	Pro	9,35	PsF	33,00	STv	28,75	PFt	27,23	PHt	31,74
CFn	6,00	CAt	16,00	CAt	18,57	PFt	9,70	OFt	34,00	Arm	29,13	CFM	27,50	CrF	33,53
CPt	7,00	PsF	16,25	PtF	19,33	PsF	10,21	CrF	35,00	Hbt	29,25	CHt	28,27	CrH	33,82
BtF	8,00	CFt	16,50	FHt	19,67	EFt	10,41	EFt	36,00	PcF	31,25	DFl	29,27	Pet	34,51
Arm	9,00	EFt	17,00	CFn	20,73	FFt	12,25	FFt	37,00	PEt	31,75	LFt	29,43	Arm	34,66
FmC	10,00	OFt	17,88	CPt	21,05	Oft	12,62	LPt	38,00	Seg	31,75	Pro	30,03	AtN	34,73
PbF	11,00	FFt	18,00	PbF	21,07	LFt	13,03	Pug	39,00	DsF	32,00	Pet	30,07	CPt	34,95
PtF	12,00	Pro	18,75	DFt	21,33	CAt	13,24	Hbt	40,00	AtN	32,25	LPt	30,63	CFn	35,04
DHt	13,00	PFt	19,00	PCF	21,40	Dfl	15,98	CFI	41,00	BtF	32,50	Pug	31,10	FmC	36,02
Seg	14,00	NSF	19,13	DHt	21,47	CPF	16,42	STv	42,00	CCl	32,63	CFI	31,73	Seg	37,40
DsF	15,00	FHt	20,25	PsF	21,80	SFt	16,44	DFl	43,00	NFA	33,38	STv	32,27	DsF	38,11
SFt	16,00	DFt	21,00	Arm	21,98	NSF	19,11	PFt	44,00	CrF	33,50	CrH	32,50	CrC	38,46
NSF	17,00	LFt	22,00	CCl	22,48	Dft	19,92	LFt	45,00	CrH	34,13	CrF	34,10	DsR	39,19
FHt	18,00	PtF	22,13	FmC	22,67	FHt	21,66	CCr	46,00	PHt	35,00	Hbt	34,13	PCF	40,07
AtN	19,00	DFl	23,00	MgC	22,73	STv	21,90	CFt	47,00	DsR	35,38	DsR	34,30	CCL	40,96
MgC	20,00	MgC	23,25	Oft	23,22	LPt	22,67	PHt	48,00	CrC	36,50	CrC	35,30	BtF	41,23
FmF	21,00	PbF	25,00	EFt	23,33	CFI	22,77	NFA	49,00	MVr	36,50	NFA	38,13	MVr	46,79
FPF	22,00	LPt	25,63	BtF	24,43	Pug	22,97	CFM	50,00	APF	38,75	PHt	39,83	CCR	48,13
DFt	23,00	FPF	26,00	FFt	24,47	MgC	26,14	Pro	51,00	CCr	47,50	CCr	47,07	APF	51,19
CAt	24,00	DHt	26,50	FmF	24,47	CPl	26,57	CSt	54,00	CSt	53,88	CSt	53,92	CSt	53,85
PEt	25,00	CPt	26,88	FPF	24,70	PtF	26,73	Esp	54,00	Esp	53,88	Esp	53,92	Esp	53,85
CPl	26,00	FmF	27,25	MVr	25,97	NFA	27,56	McE	54,00	McE	53,88	McE	53,92	McE	53,85
PFr	27,00	CFI	27,75	Seg	25,97	FPF	27,60	Pgl	54,00	Pgl	53,88	Pgl	53,92	Pgl	53,85
CHt	28,00	CHt	28,13	CFt	26,33	FmF	28,33	SSt	54,00	SSt	53,88	SSt	53,92	SSt	53,85

[1] C1: Ordem inversa da contribuição do descritor para o último eixo fatorial (O'_p), obtida a partir de análises de correspondência múltipla (AFCM) seqüenciais, em que se elimina, em cada uma, a variável com a maior contribuição para este eixo; C2: Ordem média (O_s) das contribuições do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da AFCM com todos os descritores, e da ordem O'_p já definida, tal que $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: Ordem média que pondera (O_w) diferentemente as ordens anteriormente definidas, segundo $O_w = \{[(O1 + O2 + O3)/3] + O'_p\}/2$; C4: Ordem média (Oz) dos três primeiros eixos fatoriais da análise AFCM completa (O1, O2 e O3, respectivamente), ponderada pela contribuição desses eixos, tal que $Oz = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$.

3.3.4 Descritores mínimos para *Capsicum frutescens*

Considerando-se apenas os 60 acessos de *C. frutescens* na coleção, os descritores mais importantes para a espécie, de acordo com o ordenamento de todos eles, após as análises de correspondência múltipla seqüenciais (Tabela 3.11), foram: cor do fruto imaturo (CFI), comprimento do pedúnculo (CpP), superfície da semente (SSt), comprimento da planta (CPt), segregação (Seg) e largura do fruto (LFt). Caracteres como dias para início do florescimento (DFl), persistência entre fruto/pedicelo (PtF), cor do fruto maduro (CFM), cor da antera (CAt), diâmetro da haste (DHt), número de lóculos (NLl), densidade da folha (DsF), procedência (Pro), largura da planta (LPt) e cor do filamento (CFn) são candidatos a serem descartados, pois apresentaram as maiores contribuições para o último eixo fatorial nas primeiras análises realizadas.

Alguns desses caracteres, entretanto, apresentaram contribuições elevadas nos primeiros eixos fatoriais, tais como: dias para início do florescimento (DFl), que ocupou a primeira posição no primeiro eixo fatorial; cor do fruto maduro (CFM), que foi o segundo no primeiro eixo e o primeiro nos segundo e terceiro eixos; a cor da antera (CAt), que ocupou o quinto lugar no primeiro eixo, dentre outros (Tabela 3.12). Essa contradição reforça a necessidade de se considerar, mais diretamente, na composição da lista de descritores mínimos, a contribuição das variáveis para os primeiros eixos fatoriais. Outras variáveis devem mesmo ser descartadas, pois não contribuíram para a caracterização deste grupo de acessos (Tabelas 3.11 e 3.12): espécie (Esp), pubescência da haste (PHt), pubescência da folha (PbF), cor da corola (CrC), cor da mancha da corola (CCr), macho-esterilidade (McE), pigmento do cálice (Pgl), constrição do anel do cálice (CCl), apêndice na ponta do fruto (APF), pescoço na base do fruto (PcF), mistura varietal (MVr) e cor da semente (CSt).

No caso dos acessos de *C. frutescens*, outra vez observou-se que 50% dos descritores corresponderam a mais de 90% da contribuição para o primeiro eixo fatorial (Tabela 3.12); ou seja, a contribuição acumulada no primeiro eixo, da primeira variável (DFl) até a 28ª (BtF), corresponde a mais de 94,7%. Assim, novamente adotou-se como trinta o tamanho da lista de descritores mínimos, para se comparar entre os critérios estabelecidos para a sua composição.

Tabela 3.11. Ordenamento decrescente dos descritores morfológicos (Var) para acessos de *Capsicum frutescens*, após análises de correspondência múltipla sucessivas[1] e respectivas contribuições percentuais máximas (%) para o último eixo fatorial da análise.

Análise	Var	%	Análise	Var	%	Análise	Var	%	Análise	Var	%
1	DFl	9,88	15	CrH	23,33	29	PsF	44,76	–	CpP	0,10
2	PtF	26,57	16	DsR	14,92	30	FPF	16,65	–	CFl	0,10
3	CFM	16,12	17	CPl	16,73	31	CFt	21,92	–	Esp	0,00
4	CAt	22,13	18	PEt	24,01	32	OFt	26,04	–	PHt	0,00
5	DHt	10,51	19	MgC	9,76	33	FmF	27,81	–	PbF	0,00
6	NLl	12,15	20	NSF	16,08	34	Hbt	36,55	–	CrC	0,00
7	DsF	16,80	21	AtN	11,88	35	CHt	35,42	–	CCr	0,00
8	Pro	13,64	22	Pug	28,91	36	BtF	46,52	–	McE	0,00
9	LPt	25,55	23	DFt	18,79	37	FmC	31,59	–	Pgl	0,00
10	CFn	14,49	24	FHt	22,92	38	PFr	33,20	–	CCl	0,00
11	FFt	16,21	25	EFt	21,57	39	LFt	24,95	–	APF	0,00
12	NFA	16,36	26	Arm	26,33	–	Seg	24,95	–	PcF	0,00
13	PFt	24,27	27	CrF	15,38	–	CPt	24,95	–	MVr	0,00
14	STv	30,86	28	SFt	17,28	–	SSt	24,95	–	CSt	0,00

[1] Análises em que se eliminava, em cada ciclo, a variável com a maior contribuição para o último eixo fatorial da análise anterior.

Na Tabela 3.13 é apresentado o ordenamento das variáveis, de acordo com os critérios adotados para a composição da lista de descritores mínimos para *C. frutescens*. Pelo critério C1, verificou-se uma correlação de 0,83 ($p<0,01$) entre a matriz de distâncias entre acessos obtida com todos os descritores e aquela obtida apenas com os descritores mínimos. Pelo critério C2, houve correlação de 0,89 ($p<0,01$) entre essas matrizes; pelo critério C3, a correlação foi de 0,86 ($p<0,01$); e pelo critério C4, a correlação foi de 0,92 ($p<0,01$). Logo, a vantagem ocorreu, neste caso, em favor do critério C4, que resultou na seguinte lista de descritores mínimos: cor do fruto maduro (CFM), antocianina nodal (AtN), posição do fruto (PsF), procedência (Pro), dias para início do florescimento (DFl), persistência entre fruto/pedicelo (PtF), aroma (Arm), número de flores por axila (NFA), formato da haste (FHt), secção transversal (STv), comprimento da planta (CPt), densidade de ramificação (DsR), largura da planta (LPt), dias para início da frutificação (Dft), cor da antera (CAt), posição do estigma (PEt), superfície do fruto (SFt), comprimento da haste (CHt), cor da haste (CrH), cor do filamento (CFn), número de lóculos (NLL), cor da folha (CrF), brotamento abaixo da primeira bifurcação (BtF), margem do cálice (MgC), densidade da folha (DsF), forma da corola (FmC), pungência (Pug), comprimento do fruto (CFt), pose da flor (PFr) e diâmetro da haste (DHt).

Tabela 3.12. Descritores (Var) que mais contribuíram para os três primeiros eixos fatoriais da análise de correspondência múltipla, e respectivas contribuições (%)[1] para a variabilidade nestes eixos, em ordem decrescente, para acessos de *Capsicum frutescens* do Banco de Germoplasma da Embrapa Hortaliças.

Var	Eixo 1	Var	Eixo 2	Var	Eixo 3	Var	Eixo 1	Var	Eixo 2	Var	Eixo 3
DFl	9,88	CFM	9,11	CFM	10,47	LFt	0,64	NLl	0,53	NSF	0,87
CFM	9,84	PsF	9,04	AtN	9,82	NLl	0,62	CrF	0,50	DFl	0,56
FHt	6,64	PEt	6,39	Seg	7,18	SFt	0,51	PFt	0,42	FPF	0,56
CrF	5,95	DFl	6,36	Arm	5,70	Pug	0,50	DsF	0,37	FHt	0,53
CAt	5,73	NFA	5,66	NLl	5,55	DHt	0,49	STv	0,29	SSt	0,44
STv	5,59	Pro	5,45	STv	4,62	CpP	0,45	Seg	0,24	CFt	0,42
DFt	5,09	DFt	5,40	CPl	4,42	CFI	0,45	CpP	0,15	LFt	0,37
LPt	4,47	CPt	4,73	SFt	4,36	Hbt	0,37	CFI	0,15	DFt	0,35
Pro	4,42	FHt	4,56	PtF	4,27	PFt	0,36	Hbt	0,10	CrF	0,28
NFA	3,98	AtN	4,51	PsF	4,23	CPl	0,32	FmC	0,09	EFt	0,24
AtN	3,60	SFt	4,12	FmC	3,53	NSF	0,25	SSt	0,08	OFt	0,16
CFn	3,47	DsR	3,69	MgC	3,15	PFr	0,16	CPl	0,06	FmF	0,05
PsF	3,21	Arm	3,21	LPt	2,94	SSt	0,10	FPF	0,04	CpP	0,00
PtF	3,05	PtF	2,79	DsR	2,65	Seg	0,08	EFt	0,01	Esp	0,00
CPt	2,85	CHt	2,73	Pro	2,65	OFt	0,02	MgC	0,01	PHt	0,00
MgC	2,64	PFr	2,67	DsF	2,57	Esp	0,00	Esp	0,00	Hbt	0,00
CrH	2,02	NSF	2,43	PFt	2,12	PHt	0,00	PHt	0,00	PbF	0,00
DsR	1,87	CFt	2,17	CHt	2,08	PbF	0,00	PbF	0,00	CrC	0,00
FPF	1,62	DHt	2,00	CrH	2,04	CrC	0,00	CrC	0,00	CCr	0,00
Arm	1,52	BtF	1,95	BtF	1,51	CCr	0,00	CCr	0,00	McE	0,00
CHt	1,28	Pug	1,74	PFr	1,48	McE	0,00	McE	0,00	Pgl	0,00
PEt	1,15	CAt	1,39	CPt	1,44	Pgl	0,00	Pgl	0,00	CCl	0,00
EFt	1,01	FmF	1,33	Pug	1,27	CCl	0,00	CCl	0,00	CFI	0,00
DsF	0,90	CrH	0,89	CFn	1,25	APF	0,00	APF	0,00	APF	0,00
FmC	0,88	LFt	0,75	CAt	1,00	FFt	0,00	FFt	0,00	FFt	0,00
FmF	0,69	CFn	0,70	PEt	1,00	PcF	0,00	PcF	0,00	PcF	0,00
CFt	0,67	LPt	0,64	NFA	0,97	MVr	0,00	MVr	0,00	MVr	0,00
BtF	0,66	OFt	0,55	DHt	0,90	CSt	0,00	CSt	0,00	CSt	0,00

[1] Os três eixos explicaram, juntos, 19,00% da variabilidade total captada pela análise (eixo 1: 7,22%, eixo 2: 6,04% e eixo 3: 5,74%).

Outra vez, observou-se que cerca de um terço dos descritores selecionados são caracteres relacionados ao fruto, o que corrobora os resultados aqui obtidos para as outras três espécies avaliadas e os de Teixeira (1996) e Sudré (2003). Confirmou-se também a importância da variável procedência (Pro) e a constatação de que a origem geográfica do germoplasma tem associação direta com a divergência genética entre os acessos, também em *C. frutescens*.

Tabela 3.13. Ordem crescente dos descritores (Var) usados para a caracterização de acessos de *Capsicum frutescens* do Banco de Germoplasma da Embrapa Hortaliças, obtida a partir da aplicação de quatro critérios (C1, C2, C3 e C4)[1] para escolha da lista de descritores mínimos.

C1		C2		C3		C4		C1		C2		C3		C4	
Var	O'_p	Var	O_s	Var	O_w	Var	Oz	Var	O'_p	Var	O_s	Var	O_w	Var	Oz
CFI	1,50	PsF	10,25	CPt	9,52	CFM	1,38	DsR	29,00	Pug	24,75	CrH	24,93	PFr	26,63
CpP	1,50	CFM	11,50	PsF	12,27	AtN	7,96	CrH	30,00	FmF	25,25	OFt	24,97	DHt	27,04
SSt	4,50	AtN	11,75	CHt	14,10	PsF	8,60	STv	31,00	NLl	25,75	Hbt	25,33	Seg	27,67
CPt	4,50	CPt	12,38	BtF	16,10	Pro	9,71	PFt	32,00	FPF	26,38	LPt	25,83	NSF	28,99
Seg	4,50	Arm	14,00	AtN	16,13	Dfl	10,87	NFA	33,00	DsF	27,50	EFt	26,75	PFt	29,05
LFt	4,50	CHt	16,00	Arm	16,20	PtF	12,49	FFt	34,00	NSF	27,50	DFl	26,93	CPl	29,27
PFr	7,00	FHt	16,25	Seg	16,72	Arm	12,94	CFn	35,00	CPl	28,25	NSF	27,00	FmF	29,28
FmC	8,00	Pro	16,63	FmC	16,80	NFA	13,55	LPt	36,00	PFt	29,25	CFn	27,43	FPF	29,47
BtF	9,00	SFt	16,75	FHt	16,87	FHt	13,67	Pro	37,00	SSt	29,38	PtF	27,83	LFt	29,54
CHt	10,00	DFt	18,00	LFt	16,88	STv	14,58	DsF	38,00	CFI	30,00	CAt	28,57	EFt	33,73
Hbt	11,00	DsR	18,38	PFr	16,97	CPt	14,89	NLl	39,00	CpP	30,00	CPl	29,20	Oft	37,02
FmF	12,00	NFA	18,75	SFt	17,60	DsR	15,04	DHt	40,00	DHt	30,00	PFt	30,83	SSt	37,95
OFt	13,00	STv	19,00	DFt	18,37	LPt	15,55	CAt	41,00	OFt	30,75	NLl	31,00	CFI	39,05
CFt	14,00	BtF	19,25	CrF	19,73	Dft	15,76	CFM	42,00	EFt	30,88	DsF	31,27	CPF	39,05
FPF	15,00	PEt	19,38	CFt	19,93	CAt	16,60	PtF	43,00	Hbt	33,13	DHt	33,50	Hbt	40,09
PsF	16,00	DFl	19,88	CFI	20,03	Pet	17,02	DFl	44,00	FFt	45,63	FFt	41,80	APF	49,55
SFt	17,00	PtF	20,00	CpP	20,03	SFt	17,69	Esp	50,50	APF	49,75	APF	50,05	CCL	49,55
CrF	18,00	FmC	20,50	FmF	20,37	CHt	18,19	PHt	50,50	CCl	49,75	CCl	50,05	CCR	49,55
Arm	19,00	Seg	20,88	SSt	21,35	CrH	19,83	PbF	50,50	CCr	49,75	CCr	50,05	CrC	49,55
EFt	20,00	LPt	21,00	CFM	21,70	CFn	20,08	CrC	50,50	CrC	49,75	CrC	50,05	CSt	49,55
FHt	21,00	PFr	21,00	PEt	21,80	NLL	22,13	CCr	50,50	CSt	49,75	CSt	50,05	Esp	49,55
DFt	22,00	CrF	22,25	DsR	22,03	CrF	22,23	McE	50,50	Esp	49,75	Esp	50,05	FFt	49,55
Pug	23,00	CrH	22,50	FPF	22,20	BtF	23,04	Pgl	50,50	McE	49,75	McE	50,05	McE	49,55
AtN	24,00	CFt	23,25	NFA	22,93	MgC	23,22	CCl	50,50	MVr	49,75	MVr	50,05	MVr	49,55
NSF	25,00	CAt	23,38	STv	23,00	DsF	24,13	APF	50,50	PbF	49,75	PbF	50,05	PbF	49,55
MgC	26,00	LFt	23,38	Pro	23,23	FmC	24,90	PcF	50,50	PcF	49,75	PcF	50,05	PCF	49,55
PEt	27,00	MgC	24,13	Pug	24,47	Pug	25,78	MVr	50,50	Pgl	49,75	Pgl	50,05	Pgl	49,55
CPl	28,00	CFn	24,25	MgC	24,88	CFt	26,25	CSt	50,50	PHt	49,75	PHt	50,05	PHt	49,55

[1] C1: Ordem inversa da contribuição do descritor para o último eixo fatorial (O'_p), obtida a partir de análises de correspondência múltipla (AFCM) seqüenciais, em que se elimina, em cada uma, a variável com a maior contribuição para este eixo; C2: Ordem média (O_s) das contribuições do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da AFCM com todos os descritores, e da ordem O'_p já definida, tal que $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: Ordem média que pondera (O_w) diferentemente as ordens anteriormente definidas, segundo $O_w = \{[(O1 + O2 + O3)/3] + O'_p\}/2$; C4: Ordem média (Oz) dos três primeiros eixos fatoriais da análise AFCM completa (O1, O2 e O3, respectivamente), ponderada pela contribuição desses eixos, tal que $Oz = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$.

3.3.5 Descritores mínimos para *Capsicum* spp.

O ordenamento decrescente dos descritores, segundo as suas contribuições para o último eixo fatorial nas análises de correspondência múltipla seqüenciais, envolvendo a coleção completa (893 acessos, distribuídos nas quatro espécies), é apresentado na Tabela 3.14. A variável macho-esterilidade (McE), mais uma vez, não apresentou contribuição alguma para a discriminação dos acessos, devendo, por isso, descartada. A variável espécie (Esp) foi a que mais contribuiu para o último eixo fatorial, na primeira análise (com todos os descritores), revelando-se também, a princípio, como de menor importância para a caracterização dos acessos. Em teoria, isso significa que a informação associada à variável já está contemplada por outras variáveis, que determinam eixos fatoriais de menor ordem. Neste caso isso certamente acontece, pois a identificação de espécie está relacionada principalmente com diferenças em características da flor (Teixeira, 1996). Por outro lado, essa variável também exibiu a segunda maior contribuição para o primeiro eixo fatorial, e a maior contribuição para o segundo e terceiro eixos (Tabela 3.15). Isso pôde ser verificado, em menor grau, também para as variáveis peso do fruto (PFt), com a maior contribuição no primeiro eixo; cor da corola (CrC), com a quarta maior contribuição no primeiro eixo; constrição do anel do cálice (CCl), com a segunda maior contribuição no segundo eixo; e cor da mancha da corola (CCr), com a segunda maior contribuição no terceiro eixo.

Isso ratifica o que já se discutiu anteriormente sobre a necessidade de agregar ao critério de Jolliffe (1972, 1973), informações mais diretas da contribuição das variáveis para os primeiros eixos fatoriais da análise completa (com todos os descritores). Considerando-se esse fato, foram, então, avaliados os quatro critérios de seleção de descritores mínimos, sendo que três deles contemplam, também, essa contribuição nos três primeiros eixos fatoriais.

Tabela 3.14. Ordenamento decrescente dos descritores morfológicos (Var) para acessos de *Capsicum* spp., após análises de correspondência múltipla sucessivas[1] e respectivas contribuições percentuais máximas (%) para o último eixo fatorial da análise.

Análise	Var	%	Análise	Var	%	Análise	Var	%	Análise	Var	%
1	Esp	61,84	15	CAt	15,13	29	AtN	20,61	43	FmF	39,88
2	DHt	58,19	16	FFt	23,24	30	DFl	31,51	44	SFt	24,19
3	CrC	57,74	17	PFr	25,62	31	MgC	31,86	45	PHt	33,50
4	CCl	30,52	18	CFM	33,03	32	CFI	30,70	46	FHt	25,48
5	PFt	56,68	19	CrF	28,83	33	Hbt	27,06	47	CpP	46,79
6	CCr	53,21	20	LPt	36,43	34	FPF	46,16	48	Seg	49,56
7	FmC	53,24	21	STv	25,60	35	BtF	30,89	49	DsF	43,24
8	CFt	35,43	22	DsR	29,61	36	Arm	31,11	50	PcF	27,57
9	LFt	69,85	23	Pro	31,23	37	CPl	28,69	51	APF	26,90
10	OFt	30,84	24	Pug	44,60	38	CFn	19,03	52	MVr	33,09
11	NFA	60,93	25	NLl	29,71	39	CHt	23,82	53	Pgl	40,45
12	PbF	19,00	26	NSF	27,61	40	CrH	19,10	54	SSt	50,00
13	EFt	29,52	27	CPt	19,03	41	PEt	33,14	–	CSt	50,00
14	DFt	19,43	28	PsF	14,89	42	PtF	23,28	–	McE	0,00

[1] Análises em que se eliminava, em cada ciclo, a variável com a maior contribuição para o último eixo fatorial da análise anterior.

Observando-se as contribuições dos descritores para os três primeiros eixos fatoriais da análise completa, verificou-se, novamente, que cerca de 50% daqueles com as maiores contribuições explicaram 90% ou mais da variabilidade total associada a cada eixo (Tabela 3.15). Por exemplo, a contribuição acumulada da primeira variável (PFt) até a 28ª (PFr), no primeiro eixo fatorial, foi de 92,5%. Isso ratifica a referida redundância entre os descritores, identificada também nas coleções específicas (*C. annuum*, *C. baccatum*, *C. chinense* e *C. frutescens*), e garante a possibilidade de reduzir o número de descritores necessários à caracterização dessas coleções, sem grande prejuízo à representação de suas variabilidades genéticas. A partir dessa informação, também neste caso geral, avaliaram-se os critérios de escolha dos descritores mínimos a partir de listas com trinta variáveis.

Tabela 3.15. Descritores (Var) que mais contribuíram para os três primeiros eixos fatoriais da análise de correspondência múltipla, e respectivas contribuições (%)[1] para a variabilidade nestes eixos, em ordem decrescente, para acessos de *Capsicum* spp. do Banco de Germoplasma da Embrapa Hortaliças.

Var	Eixo 1	Var	Eixo 2	Var	Eixo 3	Var	Eixo 1	Var	Eixo 2	Var	Eixo 3
PFt	6,75	Esp	10,07	Esp	15,31	DFl	1,05	DsR	0,83	FPF	0,75
Esp	6,36	CCl	7,55	CCr	14,05	AtN	1,02	SFt	0,83	DFt	0,62
NFA	5,76	EFt	5,49	CAt	11,64	Hbt	0,78	CHt	0,79	SFt	0,57
CrC	5,62	NLl	5,15	MgC	5,58	STv	0,51	FmF	0,77	CPl	0,57
LFt	5,23	OFt	5,04	CpP	4,72	NLl	0,51	MgC	0,71	CFl	0,55
NSF	4,99	FFt	4,68	Pro	3,70	PEt	0,44	Pug	0,61	CrH	0,51
CFt	4,83	LFt	4,58	PEt	3,64	CHt	0,32	AtN	0,59	Hbt	0,50
Pug	4,54	PFr	4,31	PFt	3,22	SFt	0,32	FmC	0,55	CHt	0,48
EFt	4,53	CFM	4,09	EFt	2,32	FmF	0,30	LPt	0,49	CFt	0,47
LPt	4,18	CrC	3,72	FFt	2,26	FmC	0,30	CrF	0,42	FHt	0,45
Pro	3,78	PsF	3,55	APF	2,19	CrH	0,24	DsF	0,41	PFr	0,43
OFt	3,07	NFA	3,22	LFt	2,12	CFl	0,24	Seg	0,40	CPt	0,41
DFt	3,02	PFt	3,21	NFA	2,12	CpP	0,24	CrH	0,34	DHt	0,39
BtF	2,99	FPF	2,75	FmF	1,88	DHt	0,21	PcF	0,30	DsF	0,37
CCl	2,97	PtF	2,67	PsF	1,67	CrF	0,20	CFl	0,29	PcF	0,37
MgC	2,71	STv	2,12	OFt	1,50	DsF	0,19	BtF	0,25	DFl	0,36
CPl	2,40	CCr	2,09	NLl	1,41	Seg	0,13	PHt	0,24	Seg	0,34
Arm	2,39	NSF	1,81	AtN	1,41	FHt	0,10	FHt	0,19	BtF	0,28
PsF	2,29	Arm	1,80	CFM	1,29	CCr	0,08	PEt	0,15	PbF	0,21
CFn	2,00	CAt	1,70	LPt	1,17	CSt	0,08	CpP	0,12	Pgl	0,18
CAt	1,74	DFt	1,66	STv	1,13	PcF	0,07	CPt	0,09	Arm	0,05
PtF	1,71	CPl	1,65	CrC	1,10	MVr	0,07	DHt	0,08	CCl	0,04
FFt	1,70	Hbt	1,53	FmC	1,10	PbF	0,06	Pgl	0,08	CSt	0,04
CFM	1,64	Pro	1,45	Pug	0,98	PHt	0,00	APF	0,06	SSt	0,04
CPt	1,53	CFn	1,39	CFn	0,89	McE	0,00	SSt	0,05	PHt	0,02
FPF	1,43	CFt	1,21	NSF	0,87	Pgl	0,00	MVr	0,02	MVr	0,02
DsR	1,23	PbF	0,96	PtF	0,87	APF	0,00	McE	0,00	DsR	0,01
PFr	1,15	DFl	0,89	CrF	0,83	SSt	0,00	CSt	0,00	McE	0,00

[1] Os três eixos explicaram, juntos, 12,39% da variabilidade total captada pela análise (eixo 1: 5,99%, eixo 2: 3,67% e eixo 3: 2,73%).

O ordenamento dos descritores de acordo com os quatro critérios avaliados é apresentado na Tabela 3.16. Conforme já informado, a melhor capacidade descritiva é atribuída ao descritor com a menor ordem. Então, tomando-se os trinta descritores mais bem classificados pelos critérios C1, C2, C3 e C4, observaram-se, respectivamente, as correlações de 0,80, 0,89, 0,86 e 0,87, todas significativas ($p<0,01$) pelo teste de Mantel, entre suas matrizes de distâncias entre acessos (obtidas com os respectivos descritores

mínimos) e a correspondente matriz obtida a partir de todos os descritores. Sendo assim, escolheu-se, mais uma vez, o critério C2 para a composição da lista final dos descritores mínimos.

Tabela 3.16. Ordem crescente dos descritores (Var) usados para a caracterização de acessos de *Capsicum* spp. do Banco de Germoplasma da Embrapa Hortaliças, obtida a partir da aplicação de quatro critérios (C1, C2, C3 e C4)[1] para escolha da lista de descritores mínimos.

C1		C2		C3		C4		C1		C2		C3		C4	
Var	O'_p	Var	O_s	Var	O_w	Var	Oz	Var	O'_p	Var	O_s	Var	O_w	Var	Oz
CSt	1,50	Esp	14,75	PtF	17,43	Esp	1,48	CPt	29,00	AtN	27,38	SSt	27,18	AtN	28,73
SSt	1,50	EFt	16,00	CFn	20,50	PFt	6,10	NSF	30,00	Hbt	28,00	LFt	27,33	Hbt	29,51
Pgl	3,00	LFt	17,88	CPl	20,77	EFt	7,22	NLl	31,00	PFr	28,50	CrH	27,37	BtF	29,94
MVr	4,00	NFA	18,13	CpP	21,17	LFt	7,24	Pug	32,00	CCr	29,13	AtN	27,67	FmF	30,69
APF	5,00	PFt	18,25	FmF	21,20	NFA	7,76	Pro	33,00	CFt	29,50	CFM	27,83	Pet	31,90
PCF	6,00	PsF	18,25	PsF	21,63	CrC	9,85	DsR	34,00	CHt	29,88	MVr	28,03	Dfl	32,01
DsF	7,00	Pro	18,50	MgC	21,73	Oft	10,81	STv	35,00	CCl	30,00	Esp	28,20	SFt	32,84
Seg	8,00	PtF	19,38	SFt	22,22	Pro	13,75	LPt	36,00	APF	30,50	OFt	28,37	FmC	33,75
CpP	9,00	MgC	19,50	FPF	22,40	NSF	14,07	CrF	37,00	BtF	31,25	CAt	28,43	DsR	33,91
FHt	10,00	FFt	19,75	NSF	22,73	FFt	15,10	CFM	38,00	DFl	31,75	LPt	28,83	CHt	34,28
PHt	11,00	Oft	19,75	PEt	23,07	PsF	15,75	PFr	39,00	CrH	32,75	PFt	28,93	CPF	34,66
SFt	12,00	NSF	20,13	Arm	23,30	CAt	16,74	FFt	40,00	DsF	33,13	DFl	29,33	CPt	35,41
FmF	13,00	CAt	21,25	APF	23,43	MgC	18,39	CAt	41,00	Seg	34,50	STv	29,47	CrF	38,21
PtF	14,00	NLl	21,25	Pro	23,50	CFM	18,46	DFt	42,00	CFI	35,00	PHt	30,93	CrH	38,97
Pet	15,00	CFn	22,00	DsF	24,47	CCL	19,08	EFt	43,00	FHt	35,00	DFt	31,10	CFI	39,35
CrH	16,00	CPl	22,38	EFt	25,00	Dft	19,12	PbF	44,00	PcF	35,00	CFI	31,57	DsF	42,19
CHt	17,00	CrC	22,38	NLl	25,00	Pug	19,23	NFA	45,00	CPt	35,75	PFr	31,63	PbF	43,01
CFn	18,00	CFM	22,50	CHt	25,57	CFt	19,24	OFt	46,00	FmC	36,25	CrC	31,97	Seg	43,52
CPl	19,00	FPF	22,75	PcF	25,57	LPt	20,20	LFt	47,00	DsR	36,38	CPt	33,00	APF	43,93
Arm	20,00	FmF	24,13	Seg	25,67	NLL	20,75	CFt	48,00	CrF	36,50	DsR	34,65	FHt	44,24
BtF	21,00	Pug	24,50	CSt	26,20	PtF	20,92	FmC	49,00	CSt	38,75	CFt	34,67	DHt	44,30
FPF	22,00	CpP	25,50	Hbt	26,30	CPl	21,68	CCr	50,00	Pgl	38,88	CCl	36,13	PCF	45,74
Hbt	23,00	LPt	25,75	Pug	26,47	CFn	22,58	PFt	51,00	SSt	40,00	CrF	37,17	CSt	50,64
CFI	24,00	Pet	25,75	BtF	26,77	FPF	23,11	CCl	52,00	MVr	40,25	CCr	37,60	PHt	51,22
MgC	25,00	STv	26,13	NFA	26,77	PFr	24,50	CrC	53,00	PHt	40,88	FmC	41,00	Pgl	51,64
DFl	26,00	Arm	26,50	FFt	26,93	STv	25,08	DHt	54,00	PbF	42,25	PbF	42,97	MVr	51,71
AtN	27,00	DFt	26,50	FHt	26,93	Arm	25,13	Esp	55,00	DHt	46,88	DHt	49,28	SSt	53,04
PsF	28,00	SFt	27,13	Pgl	27,12	CCR	28,44	McE	56,00	McE	55,38	McE	55,52	McE	54,88

[1] C1: Ordem inversa da contribuição do descritor para o último eixo fatorial (O'_p), obtida a partir de análises de correspondência múltipla (AFCM) seqüenciais, em que se elimina, em cada uma, a variável com a maior contribuição para este eixo; C2: Ordem média (O_s) das contribuições do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da AFCM com todos os descritores, e da ordem O'_p já definida, tal que $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: Ordem média que pondera (O_w) diferentemente as ordens anteriormente definidas, segundo $O_w = \{[(O1 + O2 + O3)/3] + O'_p\}/2$; C4: Ordem média (Oz) dos três primeiros eixos fatoriais da análise AFCM completa (O1, O2 e O3, respectivamente), ponderada pela contribuição desses eixos, tal que $Oz = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$.

A partir da aplicação desse critério (C2), os descritores mínimos, inicialmente eleitos para a caracterização dos acessos de *Capsicum* spp. da coleção em estudo, são: espécie (Esp), espessura da parede do fruto (EFt), largura do fruto (LFt), número de flores por axila (NFA), peso do fruto (PFt), posição do fruto (PsF), procedência (Pro), persistência entre fruto/pedicelo (PtF), margem do cálice (MgC), formato do fruto (FFt), ombro do fruto (OFt), número de sementes por fruto (NSF), cor da antera (CAt), número de lóculos (NLl), cor do filamento (CFn), comprimento da placenta (CPl), cor da corola (CrC), cor do fruto maduro (CFM), formato da ponta do fruto (FPF), formato da folha (FmF), pungência (Pug), comprimento do pedúnculo (CpP), largura da planta (LPt), posição do estigma (PEt), secção transversal (STv), aroma (Arm), dias para início da frutificação (DFt), superfície do fruto (SFt), antocianina nodal (AtN), hábito de crescimento (Hbt) e pose da flor (PFr). As variáveis comprimento da haste (CHt), comprimento da planta (CPt) e comprimento do fruto (CFt), embora não tenham atingido, neste caso, relevância estatística suficiente para entrar na relação anterior, merecem ser incluídos na lista de descritores mínimos de *Capsicum* spp., pois estão relacionadas nas listas já estabelecidas para as coleções das quatro espécies estudadas.

Sobre a inclusão da variável espécie (Esp) entre os descritores mínimos para acessos de *Capsicum* spp., cabe destacar, primeiramente, que se o acesso referir-se a material cultivado com espécie já identificada, a sua caracterização pode ser remetida diretamente à lista de descritores mínimos da sua espécie. Caso contrário, a inclusão dessa variável deve ser interpretada como a necessidade de manter, entre os descritores mínimos de uma coleção genérica, todos os atributos morfológicos envolvidos na chave de identificação de espécies em *Capsicum* spp. Isso, naturalmente, demonstrou-se irrelevante dentro das coleções específicas de *C. annuum*, *C. baccatum*, *C. chinense* e *C. frutescens*.

Também para a coleção completa, observou-se alta proporção dos descritores mínimos relacionados a características do fruto, corroborando os resultados obtidos dentro de espécies e os de Teixeira (1996) e Sudré (2003). Ratificou-se também a importância da variável procedência (Pro) e a constatação de que a origem geográfica do germoplasma, independentemente da espécie, tem associação direta com a divergência genética entre os acessos.

3.3.6 Síntese dos resultados

Na Tabela 3.17 são apresentadas todas as listas de descritores mínimos propostas neste estudo, seja genericamente para *Capsicum* spp. ou para as suas quatro espécies cultivadas (*C. annuum*, *C. baccatum*, *C. chinense* e *C. frutescens*). Dos 56 descritores avaliados, oito (APF, CCl, CCr, CSt, McE, MVr, PcF e SSt) mostraram-se completamente dispensáveis para a caracterização dos acessos da coleção em estudo. Alguns desses descritores, como por exemplo, constrição do anel do cálice (CCl), podem, ocasionalmente, ser considerados pelos melhoristas, pois são tidos como essenciais para a caracterização de *Capsicum*, segundo Sudré (2006). Cinco outros foram selecionados exclusivamente para uma das coleções (Esp, para *Capsicum* spp.; Pgl, para *C. annuum*; AtN, Arm e STv, para *C. frutescens*), sendo que alguns tiveram importância relativamente pequena (ex.: Pgl, STv e Seg). Esse fato evidencia, prontamente, a possibilidade de redução da lista de descritores atualmente adotada no Banco de Germoplasma de *Capsicum* da Embrapa Hortaliças.

Entre os descritores mais relevantes, dez foram selecionados para as cinco listas, a saber: cor do filamento (CFn), cor do fruto maduro (CFM), número de lóculos (NLl), posição do fruto (PsF), procedência (Pro), pungência (Pug), superfície do fruto (SFt), comprimento da haste (CHt), comprimento da planta (CPt) e comprimento do fruto (CFt). Tais descritores são, portanto, muito importantes para a caracterização dos acessos da coleção de germoplasma em estudo.

Nesse sentido, também merecem destaque outros quatorze descritores comuns para três espécies: brotamento abaixo da primeira bifurcação (BtF), cor da antera (CAt), cor do fruto imaturo (CFI), dias para início do florescimento (DFl), dias para início da frutificação (DFt), espessura da parede do fruto (EFt), formato do fruto (FFt), formato da ponta do fruto (FPF), largura da planta (LPt), margem do cálice (MgC), número de sementes por fruto (NSF), ombro do fruto (OFt), pose da flor (PFr) e peso do fruto (PFt). Destes, apenas três não se repetem na lista genérica de *Capsicum* spp. (BtF, CFI e DFl). Outros quinze descritores mínimos são comuns a duas espécies, e oito deles (CPl, CpP, CrC, FmF, Hbt, NFA, PtF e PEt) estão também na lista genérica. Logo, os descritores mínimos selecionados para a coleção completa (*Capsicum* spp.) estão, em grande parte, representados entre aqueles das coleções específicas. Apesar disso, segundo Valls (2007), de modo geral, os descritores mais adequados para o discernimento entre espécies de um gênero tendem a ser os menos adequados para a diferenciação entre cultivares de uma única espécie.

Tabela 3.17. Listas de descritores mínimos propostas para a caracterização de *Capsicum* spp. e de suas espécies cultivadas *C. annuum*, *C. baccatum*, *C. chinense* e *C. frutescens*[1], obtidas a partir da análise de 893 acessos do Banco de Germoplasma da Embrapa Hortaliças, caracterizados segundo 56 descritores morfológicos[2].

Capsicum spp. (34)	*C. annuum* (30)	*C. baccatum* (30)	*C. chinense* (32)	*C. frutescens* (30)
Espécie (Esp)	Comprimento do fruto (CFt)	Formato da ponta do fruto (FPF)	Pose da flor (PFr)	Cor do fruto maduro (CFM)
Espessura da parede do fruto (EFt)	Pigmento do cálice (Pgl)	Posição do fruto (PsF)	Número de lóculos (NLl)	Antocianina nodal (AtN)
Largura do fruto (LFt)	Espessura da parede do fruto (EFt)	Comprimento do fruto (CFt)	Comprimento do pendúnculo (CpP)	Posição do fruto (PsF)
Número de flores por axila (NFA)	Largura do fruto (LFt)	Peso do fruto (PFt)	Cor do fruto maduro (CFM)	Procedência (Pro)
Peso do fruto (PFt)	Procedência (Pro)	Formato do fruto (FFt)	Superfície do fruto (SFt)	Dias para início do florescimento (DFl)
Posição do fruto (PsF)	Peso do fruto (PFt)	Espessura da parede do fruto (EFt)	Cor da antera (CAt)	Persistência entre fruto/pedicelo (PtF)
Procedência (Pro)	Cor da haste (CrH)	Superfície do fruto (SFt)	Posição do fruto (PsF)	Aroma (Arm)
Persistência entre fruto/pedicelo (PtF)	Cor da corola (CrC)	Cor da corola (CrC)	Comprimento do fruto (CFt)	Número de flores por axila (NFA)
Margem do cálice (MgC)	Cor do fruto imaturo (CFI)	Comprimento da haste (CHt)	Espessura da parede do fruto (EFt)	Formato da haste(FHt)
Formato do fruto (FFt)	Hábito de crescimento (Hbt)	Comprimento da planta (CPt)	Ombro do fruto (OFt)	Secção transversal (STv)
Ombro do fruto (OFt)	Cor do filamento (CFn)	Número de lóculos (NLl)	Formato do fruto (FFt)	Comprimento da planta (CPt)
Número de sementes por fruto (NSF)	Comprimento da haste (CHt)	Cor do fruto imaturo (CFI)	Procedência (Pro)	Densidade de ramificação (DsR)
Cor da antera (CAt)	Cor da antera (CAt)	Procedência (Pro)	Peso do fruto (PFt)	Largura da planta (LPt)
Número de lóculos (NLl)	Ombro do fruto (OFt)	Comprimento do pendúnculo (CpP)	Número de sementes por fruto (NSF)	Dias para início da frutificação (DFt)
Cor do filamento (CFn)	Formato da ponta do fruto (FPF)	Pungência (Pug)	Formato da haste (FHt)	Cor da antera (CAt)
Comprimento da placenta (CPl)	Brot. abaixo da 1ª bifurcação (BtF)	Dias para início do florescimento (DFl)	Dias para início da frutificação (DFt)	Posição do estigma (PEt)
Cor da corola (CrC)	Posição do fruto (PsF)	Número de flores por axila (NFA)	Largura do fruto (LFt)	Superfície do fruto (SFt)
Cor do fruto maduro (CFM)	Densidade da folha (DsF)	Brot. abaixo da 1ª bifurcação (BtF)	Persistência entre fruto/pedicelo (PtF)	Comprimento da haste (CHt)
Formato da ponta do fruto (FPF)	Cor da folha (CrF)	Cor do filamento (CFn)	Dias para inicio do florescimento (DFl)	Cor da haste (CrH)
Formato da folha (FmF)	Formato do fruto (FFt)	Número de sementes por fruto (NSF)	Margem do cálice (MgC)	Cor do filamento (CFn)
Pungência (Pug)	Número de sementes por fruto (NSF)	Largura do fruto (LFt)	Pubescência da folha (PbF)	Número de lóculos (NLl)
Comprimento do pendúnculo (CpP)	Pungência (Pug)	Pubescência da folha (PbF)	Largura da planta (LPt)	Cor da folha (CrF)
Largura da planta (LPt)	Superfície do fruto (SFt)	Posição do estigma (PEt)	Formato da ponta do fruto (FPF)	Brotamento à primeira bifurcação (BtF)
Posição do estigma (PEt)	Margem do cálice (MgC)	Pubescência da haste (PHt)	Diâmetro da haste (DHt)	Margem do cálice (MgC)
Secção transversal (STv)	Cor do fruto maduro (CFM)	Ombro do fruto (OFt)	Comprimento da planta (CPt)	Densidade da folha (DsF)
Aroma (Arm)	Comprimento da planta (CPt)	Comprimento da placenta (CPl)	Formato da folha (FmF)	Forma da corola (FmC)
Dias para início da frutificação (DFt)	Pose da flor (PFr)	Cor do fruto maduro (CFM)	Cor do fruto imaturo (CFI)	Pungência (Pug)
Superfície do fruto (SFt)	Número de lóculos (NLl)	Dias para início da frutificação (DFt)	Comprimento da haste (CHt)	Comprimento do fruto (CFt)
Antocianina nodal (AtN)	Pubescência da haste (PHt)	Densidade de ramificação (DsR)	Cor do filamento (CFn)	Pose da flor (PFr)
Hábito de crescimento (Hbt)	Largura da planta (LPt)	Hábito de crescimento (Hbt)	Comprimento da placenta (CPl)	Diâmetro da haste (DHt)
Pose da flor (PFr)			Forma da corola (FmC)	
Comprimento do fruto (CFt)			Pungência (Pug)	
Comprimento da haste (CHt)				
Comprimento da planta (CPt)				

[1] O valor entre parênteses no cabeçalho representa o número de descritores em cada lista.
[2] A ordem de apresentação das variáveis, em cada lista, tem relação com a importância do descritor para a caracterização do germoplasma correspondente.

3.4 CONCLUSÕES

- A análise fatorial de correspondência múltipla demonstra ser apropriada para a seleção de descritores mínimos, em análise multivariada de dados qualitativos, provenientes de coleções de germoplasma.
- O critério de descarte de variáveis baseado na proposta de Jolliffe (1972, 1973) mostra ser menos eficiente do que alternativas que valorizam mais a contribuição das variáveis para os primeiros eixos fatoriais.
- Existe considerável grau de associação entre os descritores inicialmente considerados na caracterização do germoplasma, de modo que é possível reduzi-los consideravelmente, sem grande prejuízo à representação da variabilidade genética. As reduções propostas implicam no descarte de quase 50% desses descritores.
- Dez descritores mínimos mostram-se relevantes para as quatro espécies estudadas: cor do filamento (CFn), cor do fruto maduro (CFM), número de lóculos (NLl), posição do fruto (PsF), procedência (Pro), pungência (Pug), superfície do fruto (SFt), comprimento da haste (CHt), comprimento da planta (CPt) e comprimento do fruto (CFt).
- Oito dos descritores avaliados mostraram-se completamente dispensáveis para a caracterização dos acessos da coleção estudada: apêndice na ponta do fruto (APF), constrição do anel do cálice (CCl), cor da mancha da corola (CCr), cor da semente (CSt), macho-esterilidade (McE), mistura varietal (MVr), pescoço na base do fruto (PcF) e superfície da semente (SSt).
- Os descritores mínimos selecionados, em grande parte, são relacionados a características do fruto. Dada a importância desses caracteres para o melhoramento genético de *Capsicum*, demonstra-se que é possível caracterizar coleções de germoplasma de modo a atender, simultaneamente, aos critérios de representação da variabilidade genética, assim como aos interesses mais imediatos do melhorista, o que aumenta as possibilidades de utilização do germoplasma.
- A relevância atribuída à variável procedência dos acessos (Pro), em todas as análises, revela que a origem geográfica do germoplasma, independentemente da espécie, tem associação direta com a divergência genética entre os acessos em *Capsicum* spp.

4 ESTABELECIMENTO DE COLEÇÕES NUCLEARES PARA *Capsicum* spp. A PARTIR DE ANÁLISES DE DADOS QUALITATIVOS

RESUMO

O presente estudo teve como objetivo propor, a partir da caracterização do banco de germoplasma de pimentas e pimentões da Embrapa Hortaliças, coleções nucleares para *Capsicum* spp. e suas espécies cultivadas *C. annuum*, *C. baccatum*, *C. chinense* e *C. frutescens*. Da coleção de cerca de 2.500 acessos, foram considerados 893, que estão distribuídos entre as quatro espécies e caracterizados segundo 56 descritores morfológicos de natureza qualitativa. Foram utilizadas as análises de correspondência simples, de agrupamento pelos métodos do vizinho mais próximo e de Tocher (com critério de aglomeração inverso), a partir das matrizes de distâncias genéticas para variáveis multicategóricas. As análises utilizadas foram eficazes na identificação da variabilidade genética entre os acessos e, também, de possíveis duplicidades na coleção, permitindo a escolha de acessos divergentes para a composição das coleções nucleares. Intensidades de seleção de 10% a 30% foram utilizadas para compor as coleções nucleares, conforme o tamanho das coleções iniciais. Para avaliar a qualidade das coleções nucleares relativamente às coleções base (validação), utilizaram-se índices de similaridade e de diversidade fenotípica. Todos os métodos avaliados mostraram-se eficazes no estabelecimento de coleções para *Capsicum* spp., com vantagem para o método de agrupamento hierárquico pelo vizinho mais próximo. O método de Tocher invertido, em geral, produziu coleções nucleares com menor similaridade em relação às respectivas coleções base e com menor diversidade fenotípica, comparativamente aos outros métodos. A análise de correspondência simples apresentou-se como alternativa útil para o estabelecimento de coleções nucleares a partir de bancos de germoplasma cujos acessos encontram-se caracterizadas, predominantemente, segundo descritores qualitativos. As coleções nucleares obtidas pela análise de correspondência e pelo método de agrupamento hierárquico do vizinho mais próximo foram as que apresentaram maior concordância entre si. Os pares de acessos 333 e 2912, 2957 e 2958, 3851 e 3853, 3767 e 3768, 3467 e 3485, 3241 e 3242 devido à alta similaridade, possivelmente, tratam-se de duplicatas. A coleção nuclear proposta para *Capsicum* spp., embora com apenas 17% dos acessos, possui índice de similaridade superior a 95% em relação à coleção completa, bem como índice de diversidade fenotípica superior à coleção base.

Palavras-chave: pimenta, dados qualitativos, análise de agrupamento, análise de correspondência simples.

ABSTRACT

CORE COLLECTIONS FOR *Capsicum* spp. PROPOSED FROM QUALITATIVE DATA ANALYSIS

The objective of this study was to propose core collections to *Capsicum* spp. and their cultivated species *C. Annuum*, *C. Baccatum*, *C. Chinense* and *C. Frutescens*. The data used were obtained from the characterization of the germplasm bank of sweet and chilli peppers of Embrapa Hortaliças. From the collection of around 2,500 accessions, 893were considered, and are distributed among the four species and characterized according to 56 morphological qualitative descriptors. We used the correspondence analysis, the clustering methods for the nearest neighbor and Tocher (with inverse grouping criteria), from the genetic distances matrices for multicategoric variables. The analyses used were effective in identifying the genetic variability among accessions and also possible duplications in the collection, allowing the choice of differing accessions to the composition of the core collections. Selection intensities from 10% to 30% were used to compose the core collections, according to size of the initial collections. To assess the quality of the core collections regarding the respective base collections (validation), indexes of similarity and phenotypic diversity were used. All methods evaluated showed effectiveness in establishing collections for *Capsicum* spp. by identifying a clear advantage in favor of the hierarchical clustering by nearest neighbor method. The inverted Tocher method, in general, produced core collections of smaller similarity with their respective base collections and with lower phenotypic diversity in relation to the other methods. The correspondence analysis showed to be a useful alternative to the establishment of core collections from germplasm banks whose accessions are characterized, predominantly, according to qualitative descriptors. The core collections obtained by the correspondence analysis and by the hierarchical clustering of the nearest neighbor method showed the highest agreement (common access). The pairs of accessions 333 and 2912, 2957 and 2958, 3851 and 3853, 3767 and 3768, 3467 and 3485, 3241 and 3242, due to their high similarity, possibly, can be considered duplicates. The proposed core collection to *Capsicum* spp., although with only 17% of accessions, has a similarity index higher than 95% in relation to the whole collection, as well as phenotypic diversity index above the collection base.

Key words: chilli pepper, qualitative data, cluster analysis, correspondence analysis.

4.1 INTRODUÇÃO

A imensa diversidade genética existente na natureza e seu uso potencial em plantas cultivadas resultaram no estabelecimento dos chamados centros de conservação de germoplasma, distribuídos pelo mundo. Houve grande esforço na coleta e conservação do germoplasma das principais espécies cultivadas, resultando também no estabelecimento de grandes coleções. Segundo Vasconcelos et al. (2007), cerca de 6,1 milhões de acessos de plantas são conservados em todo o mundo e esses acessos encontram-se dispostos em 1.320 bancos de germoplasma. Esses bancos estão distribuídos pelos 157 países que compõem a Comissão de Recursos Fitogenéticos da FAO. No Brasil, existem mais de 200 mil acessos de plantas em conservação *ex situ*, nos cerca de 160 bancos de germoplasma implantados em mais de cinqüenta locais. De modo geral, entretanto, o crescimento dessas coleções não foi acompanhado por equivalente intensidade de uso, o que tem gerado lacunas entre a disponibilidade do germoplasma e a real utilização desses materiais.

Segundo Basigalup et al. (1995), muitas coleções de germoplasma tornaram-se tão grandes e difusas, que desencorajaram a efetiva avaliação dos acessos e, conseqüentemente, tem atrasado a utilização do germoplasma. O excessivo tamanho dessas coleções e a limitação de recursos financeiros foram os principais fatores que contribuíram para o estabelecimento da defasagem entre o crescimento e o uso das coleções. Essa combinação de fatores tem proporcionado baixa qualidade da informação, uma vez que, com o aumento das coleções, estas não são bem caracterizadas para atributos de importância econômica.

Para resolver esse problema, o desafio é incrementar o uso do germoplasma conservado, pois, atualmente, apenas cerca de 4% dos genótipos armazenados vêm sendo utilizados no mundo (Vasconcelos et al., 2007). Por isso, é necessário priorizar a caracterização, a avaliação, a documentação e a informação sobre os acessos. Nesse sentido, muitos autores têm discutido formas de reduzir o tamanho das coleções de germoplasma, sem perdas significativas na sua variabilidade genética. Frank & Brown (1984) desenvolveram a base teórica dessa abordagem e propuseram diretivas de ordem prática para viabilizar a idéia inicialmente apresentada por Harlan & de Wet (1972), como coleções de trabalho. Sugeriram, então, o que denominaram de coleções nucleares (*core collections* ou *core subsets*, em inglês), isto é, um subconjunto representativo da coleção total, com mínimo grau de redundância entre os acessos. Segundo eles, os acessos não

incluídos na coleção nuclear devem ser preservados como um subconjunto reserva para necessidades futuras relacionadas à busca de alelos raros. A partir daí, a organização das coleções de germoplasma, em torno de coleções nucleares, começou a ganhar espaço como estratégia para racionalizar esforços na aquisição de informações acerca da variabilidade do germoplasma preservado.

Com o crescimento rápido e difuso das coleções (Basigalup et al., 1995), muitas vezes, amostras obtidas em determinada região, embora diferentes, podem receber o mesmo nome, ou amostras idênticas geneticamente, provenientes de regiões distintas, recebem identificações diferentes. Isso caracteriza as chamadas duplicidades de acessos (duplicatas), que aumentam o trabalho nos bancos de germoplasma, reduzem o espaço disponível para a conservação de outras amostras e não contribuem para o enriquecimento da sua variabilidade genética. Assim, a identificação de materiais duplicados nos bancos de recursos genéticos é tarefa importante para se aumentar a eficiência e a eficácia nos trabalhos de conservação e avaliação de germoplasma. Assim, é possível priorizar e concentrar esforços na caracterização e avaliação das coleções, de modo a formar uma base de informação mais completa sobre o conjunto de acessos, levando, efetivamente, a uma ampliação das possibilidades de uso do germoplasma.

Grande importância tem sido dada, atualmente, à obtenção de coleções nucleares, existindo tais coleções em arroz (Abadie et al., 2005; Borba, 2007), batata (Chandra et al., 2002), cevada (Liu et al., 1999), entre outras. A escolha de genótipos para compor uma coleção nuclear utiliza estratégias que envolvem diversas intensidades e métodos de amostragem (Hu et al., 2000; Thompson, 2002; Wang et al., 2007; Vasconcelos et al., 2007). A intensidade de amostragem refere-se à porcentagem de acessos da coleção original de germoplasma que será selecionada para compor a coleção nuclear. Na literatura, em geral, recomendam-se intensidades de amostragem que variam entre 5% e 25% do número total de acessos (Brown, 1989; Basigalup et al., 1995; Brown & Spillane, 1999; van Hintum et al., 2000; Chandra et al., 2002). Mas, intensidades de até 30% também são utilizadas (Yonezawa et al., 1995; Xu et al., 2004), dependendo do grau da redundância genética entre os acessos, dos recursos disponíveis para manutenção da coleção nuclear e da freqüência com que se realiza a regeneração dos acessos. Por outro lado, Upadhyaya & Ortiz (2001) utilizaram intensidade de amostragem tão baixa quanto 1% dos acessos da coleção inicial, em grão-de-bico. Esta intensidade, contudo, somente garante resultados satisfatórios no caso de coleções bastante grandes, com mais de 2.500

acessos (Vasconcelos et al., 2007). Assim, idealmente, a coleção nuclear deveria ser criada com base na caracterização e avaliação completa dos materiais, porém, geralmente esta é criada a partir de informações morfogeográficas, validadas por análise genética (Rodrigues et al., 2005).

Segundo van Hintum et al. (2000), a criação de uma coleção nuclear pode ser muito simples. Esta pode ser definida para qualquer coleção de germoplasma, mesmo que não haja uma documentação completa ou dados completamente confiáveis. É preciso existir, somente, o conhecimento básico sobre a coleção e a espécie envolvida, e algum tempo para a escolha dos acessos. Entretanto, simples ou complexo, é importante assegurar perda mínima da diversidade genética (Crossa & Franco, 2004), e que haja um trabalho conjunto entre o curador do germoplasma, os melhoristas de plantas e outros usuários interessados na cultura e no uso da sua diversidade genética.

O objetivo deste trabalho foi propor coleções nucleares para o banco de germoplasma de pimentas e pimentões mantido pela Embrapa Hortaliças, com ênfase na coleção completa de *Capsicum* spp. e nas sub-coleções associadas às quatro espécies cultivadas do gênero (*C. annuum*, *C. baccatum*, *C. chinense* e *C. frutescens*). Buscou-se, também, avaliar métodos alternativos para a proposição dessas coleções, incluindo-se procedimentos amplamente utilizados, como os métodos do vizinho mais próximo e de Tocher, além da análise fatorial de correspondências, ainda não descrita nesse tipo de aplicação.

4.2 MATERIAL E MÉTODOS

4.2.1 Origem dos dados

Os dados utilizados para o desenvolvimento deste trabalho foram obtidos junto à Embrapa Hortaliças, que possui uma coleção de germoplasma de pimentas e pimentões (*Capsicum* spp.), atualmente com cerca de 2.500 acessos. Desse total, foram considerados 893 acessos, distribuídos entre as quatro espécies cultivadas do gênero *Capsicum* (420 acessos de *C. annuum*, 106 de *C. baccatum*, 307 de *C. chinense* e 60 de *C. frutescens*). Os acessos são oriundos de diferentes regiões do país e do exterior, e se encontram completamente caracterizados segundo 56 descritores morfológicos atualmente definidos

(Tabela 4.1). As categorias de cada descritor são apresentadas em detalhe no Apêndice A. Embora a maior parte deles sejam naturalmente qualitativos, observam-se alguns descritores de natureza quantitativa (ex.: comprimento da planta, diâmetro da haste, comprimento, largura e peso do fruto, entre outros), os quais foram também categorizados com o objetivo de reuni-los todos num mesmo tipo de análise.

Para a avaliação do germoplasma, os acessos foram semeados em bandeja de isopor e, após quarenta dias, cinco mudas de cada acesso foram transplantadas em telado, no espaçamento de 1,0 m x 0,60 m. As plantas foram conduzidas de acordo com as práticas agronômicas convencionais utilizadas para a cultura (Reifschneider, 2000), tendo sido tutoradas por estacas e irrigadas por gotejamento.

Tabela 4.1. Lista dos descritores para germoplasma de *Capsicum* spp., com seu número de categorias (*K*), atualmente adotados pela Embrapa Hortaliças, com respectivas abreviaturas.

Descritores	*K*	Abrev.	Descritores	*K*	Abrev.
Antocianina nodal	4	AtN	Formato da haste	3	FHt
Apêndice na ponta do fruto	2	APF	Formato da ponta do fruto	4	FPF
Aroma	3	Arm	Formato do fruto	6	FFt
Brotamento abaixo da primeira bifurcação	4	BtF	Hábito de crescimento	4	Hbt
Comprimento da haste	4	CHt	Largura da planta	4	LPt
Comprimento da placenta	3	CPl	Largura do fruto	5	LFt
Comprimento da planta	5	CPt	Macho-esterilidade	2	McE
Comprimento do fruto	6	CFt	Margem do cálice	3	MgC
Comprimento do pedúnculo	4	CpP	Mistura varietal	2	MVr
Constrição do anel do cálice	2	CCl	Número de flores por axila	8	NFA
Cor da antera	6	CAt	Número de lóculos	5	NLl
Cor da corola	11	CrC	Número de sementes por fruto	3	NSF
Cor da folha	8	CrF	Ombro do fruto	5	OFt
Cor da haste	3	CrH	Persistência entre fruto/pedicelo	3	PtF
Cor da mancha da corola	6	CCr	Pescoço na base do fruto	2	PcF
Cor da semente	4	CSt	Peso do fruto	6	PFt
Cor do filamento	7	CFn	Pigmento do cálice	2	Pgl
Cor do fruto imaturo	10	CFI	Pose da flor	7	PFr
Cor do fruto maduro	14	CFM	Posição do estigma	6	PEt
Densidade da folha	3	DsF	Posição do fruto	7	PsF
Densidade de ramificação	3	DsR	Procedência	12	Pro
Diâmetro da haste	4	DHt	Pubescência da folha	3	PbF
Dias para início do florescimento	4	DFl	Pubescência da haste	3	PHt
Dias para início da frutificação	4	DFt	Pungência	4	Pug
Espécie	4	Esp	Secção transversal	3	STv
Espessura da parede do fruto	6	EFt	Segregação	7	Seg
Forma da corola	3	FmC	Superfície da semente	3	SSt
Formato da folha	3	FmF	Superfície do fruto	5	SFt

4.2.2 Análises estatísticas

Para a proposição das coleções nucleares foi utilizada a análise de correspondência simples, que permite explorar estatisticamente um conjunto de dados multivariados, proveniente de *n* indivíduos avaliados segundo *p* variáveis categóricas ou qualitativas. Na implementação dessa análise foi utilizado o procedimento CORRESP do sistema computacional SAS (SAS Institute, 2002), apresentado no Apêndice C.

Utilizou-se também a análise da matriz de dissimilaridade genética para variáveis multicategóricas, proposta por Cruz & Carneiro (2003), cujo índice é obtido por $d_{ii'} = D/(C+D)$, em que C é o número de concordâncias de categoria e D é o número de discordâncias de categoria. Essa matriz de distâncias foi associada ao método de agrupamento hierárquico do vizinho mais próximo, que identifica os acessos mais similares, os quais são reunidos formando o grupo inicial. Posteriormente, determinam-se as distâncias desse grupo em relação aos demais acessos, ou em relação a outros grupos formados similarmente ao grupo inicial. O processo de identificação dos grupos ou acessos mais similares se repete sobre uma nova matriz de dissimilaridade, cuja dimensão é reduzida a cada passo e finaliza quando todos os acessos são reunidos em um único grupo. Nesse método, um dendrograma é estabelecido a partir dos genótipos com maior similaridade, sendo que a distância entre um acesso *k* e um grupo formado pelos acessos *i* e *j* é dada por: $d_{(ij)k} = \min\{d_{ik}; d_{jk}\}$, ou seja, $d_{(ij)k}$ é dada pelo menor elemento do conjunto das distâncias entre pares de acessos (*i* e *k*) e (*j* e *k*). A distância entre dois grupos é dada por: $d_{(ij)(kl)} = \min\{d_{ik}; d_{il}; d_{jk}; d_{jl}\}$, ou seja, a distância entre dois grupos formados pelos acessos (*i* e *j*) e (*k* e *l*) é dada pelo menor elemento do conjunto, cujos elementos são as distâncias entre os pares de acessos (*i* e *k*), (*i* e *l*), (*j* e *k*) e (*j* e *l*). Esse método foi implementado por meio do programa computacional NTSys (Rohlf, 1989).

Associado à matriz de dissimilaridade genética para variáveis multicategóricas, utilizou-se também o método de agrupamento de Tocher (Rao, 1952), com critério de aglomeração inverso, que consiste em agrupar acessos com maior dissimilaridade. Por esse método, descrito detalhadamente em Vasconcelos et al. (2007), os acessos agrupados até um nível estabelecido de dissimilaridade formam um único grupo, que corresponde diretamente à coleção nuclear de interesse. Em resumo, encontra-se o par de indivíduos mais dissimilares, que forma o grupo inicial. Para a inclusão de um outro acesso, são calculadas as distâncias entre o grupo inicial e os demais genótipos. O acesso que

apresentar maior distância em relação a esse grupo será o genótipo incluído. Assim, se mantém a regra, até que se obtenha o grupo único com o número de acessos necessários para a formação da coleção nuclear. Este número foi definido pela intensidade de amostragem estabelecida para cada coleção. Para implementação deste método, utilizou-se o programa computacional GENES (Cruz, 2001).

Para o estabelecimento das coleções nucleares, foram adotadas intensidades de amostragem que variaram de 10% a 30%, conforme o tamanho da coleção inicial, considerando-se também a atual capacidade de trabalho da instituição responsável pelo banco de germoplasma (Embrapa Hortaliças), que ficou estabelecida em até 150 acessos para a coleção nuclear completa de *Capsicum* spp. Assim, para *C. annuum* (coleção de 420 acessos) adotou-se a intensidade de 10%; para *C. chinense* (coleção de 307 acessos), 20%; para *C. baccatum* (coleção de 106 acessos), 25%; e para *C. frutescens* (coleção de 60 acessos), 30%.

A princípio, a coleção completa de *Capsicum* spp. foi divida em quatro grupos ou estratos principais, correspondentes às referidas espécies cultivadas, seguindo-se a aplicação, dentro de cada estrato, das análises para o estabelecimento de suas coleções nucleares. No caso das análises de correspondência simples e de agrupamento pelo método do vizinho mais próximo, para cada coleção de espécie, foram constituídos grupos de similaridade, dentro dos quais aplicou-se amostragem aleatória estratificada proporcional para compor a respectiva coleção nuclear (acessos divergentes, oriundos de todos os grupos de similaridade identificados). Logo, nessa amostragem, a proporcionalidade do tamanho do grupo somente foi considerada a partir do momento que se garantiu, pelo menos, um acesso por grupo. Assim, os grupos maiores, em geral, tiveram menos acessos na coleção nuclear do que o previsto pela proporção de seus tamanhos relativamente ao tamanho desta coleção; e os grupos menores (com apenas um ou dois acessos), o contrário. Dessa forma, o corte do dendrograma obtido pelo método do vizinho mais próximo foi realizado de modo a obter, preferencialmente, tantos grupos quantos forem os acessos definidos pela intensidade de amostragem. Na análise de agrupamento pelo método de Tocher com critério de aglomeração inverso, a coleção nuclear foi obtida de forma direta, também dentro de cada coleção de espécie, porém, sem a necessidade de amostragem aleatória de acessos.

Para avaliar a qualidade das coleções nucleares constituídas segundo cada método (análise de correspondência, agrupamento pelo vizinho mais próximo e método de

Tocher), processo também chamado de validação das coleções propostas, empregaram-se duas estatísticas de caráter exploratório, isto é, não inferencial. A primeira delas foi a taxa de coincidência (CR%) proposta por Hu et al. (2000), com uma adaptação implementada neste trabalho para dados de natureza qualitativa: $CR\% = \frac{1}{m}\left(\sum_{j=1}^{m}\frac{C_j^N}{C_j^B}\right)x100$, em que m é o número de variáveis ou descritores (j = 1, 2, ..., m); e C_j é o número de categorias associadas à variável j, manifestadas na coleção nuclear (C_j^N) e na coleção base (C_j^B). Esse índice mede, portanto, o grau de concordância das categorias de variáveis na coleção nuclear e na coleção base a ser representada; ou seja, tem relação direta com o propósito de preservação, na coleção nuclear, da riqueza alélica responsável pela expressão das categorias (fenótipos) que se manifestam na coleção base.

A outra estatística empregada para validar as coleções nucleares, originalmente proposta neste trabalho, foi o coeficiente de diversidade fenotípica (DF), dado por: $DF = \sum_{j=1}^{m} D_j^N / \sum_{j=1}^{m} D_j^B$, em que $D_j = (1 - \sum_{k=1}^{C_j} f_k^2) / (1 - 1/C_j)$ é o índice de diversidade associado à variável j, descrito por Souza & Peñaloza (2005), com $C_j \geq 2$ categorias e freqüências relativas f_k (k = 1, 2, ..., C_j categorias). O índice D_j tem variação entre zero (todas as observações ou acessos na mesma categoria da variável) e um (observações igualmente distribuídas nas C_j categorias da variável), sendo acumulado para as "m" variáveis, na coleção nuclear (N) e na coleção base (B), para gerar o coeficiente DF, com variação entre zero e $p / \sum_j D_j^B$. Assim, valores DF < 1,0 indicam menor diversidade fenotípica na coleção nuclear, relativamente à coleção inicial; DF = 1,0, igual distribuição da diversidade fenotípica nas duas coleções; e valores DF > 1,0 indicam que a distribuição dos acessos nas categorias das variáveis é mais equilibrada na coleção nuclear do que na coleção base. Vale ressaltar que a riqueza de categorias não pode ser maior na coleção nuclear do que na coleção base. Assim, o coeficiente DF complementa a avaliação de CR%, pois avalia não só a representação das categorias de variáveis na coleção nuclear (riqueza), mas também o quanto a distribuição dos acessos nessas categorias é equilibrada, de modo a minimizar a probabilidade de erosão genética.

A coleção nuclear proposta para representar os acessos de cada espécie foi aquela constituída a partir do método que gerou os maiores coeficientes de similaridade (CR%) e de diversidade fenotípica (DF). Por fim, a coleção nuclear para *Capsicum* spp.,

isto é, aquela proposta para representar todos os acessos já caracterizados da coleção completa (893 acessos), foi constituída reunindo-se as coleções nucleares estabelecidas separadamente por espécie.

4.3 RESULTADOS E DISCUSSÃO

4.3.1 Coleção nuclear para *Capsicum annuum*

A análise de correspondência para os dados da coleção de *C. annuum* (420 acessos) permitiu verificar que os três primeiros eixos fatoriais da análise explicaram 46% da variabilidade dos dados. Observando-se a dispersão gráfica mostrada na Figura 4.1, nota-se a formação de três nuvens de pontos mais compactas (grupos 1, 2 e 3), em que se encontram a maior parte dos acessos da espécie. Verificou-se, também, a formação de duas nuvens menores (grupos 4 e 5), próximas aos grupos 2 e 3. O restante dos acessos se dispersaram mais entre si no grupo 6, sendo que o acesso 3745 foi o que mais se afastou na dispersão de pontos, devendo apresentar características bastante divergentes em relação aos demais.

Na composição da coleção nuclear segundo este procedimento, amostraram-se aleatoriamente acessos nos grupos ou estratos observados, visando melhor representar a variabilidade da coleção inicial. O par de acessos mais divergentes foi 3745 e 2887, que apresentam diferenças em 29 descritores, incluindo comprimento, largura e peso do fruto. O acesso 2887, representado na parte inferior do gráfico, possui frutos grandes, com peso acima de 81 g, enquanto o descritor 3745, na parte superior do gráfico, possui frutos pequenos com no máximo 1,0 g. Isso mostra que o tamanho do fruto tem grande importância na discriminação dos acessos dessa espécie, revelada nos primeiros eixos fatoriais da análise.

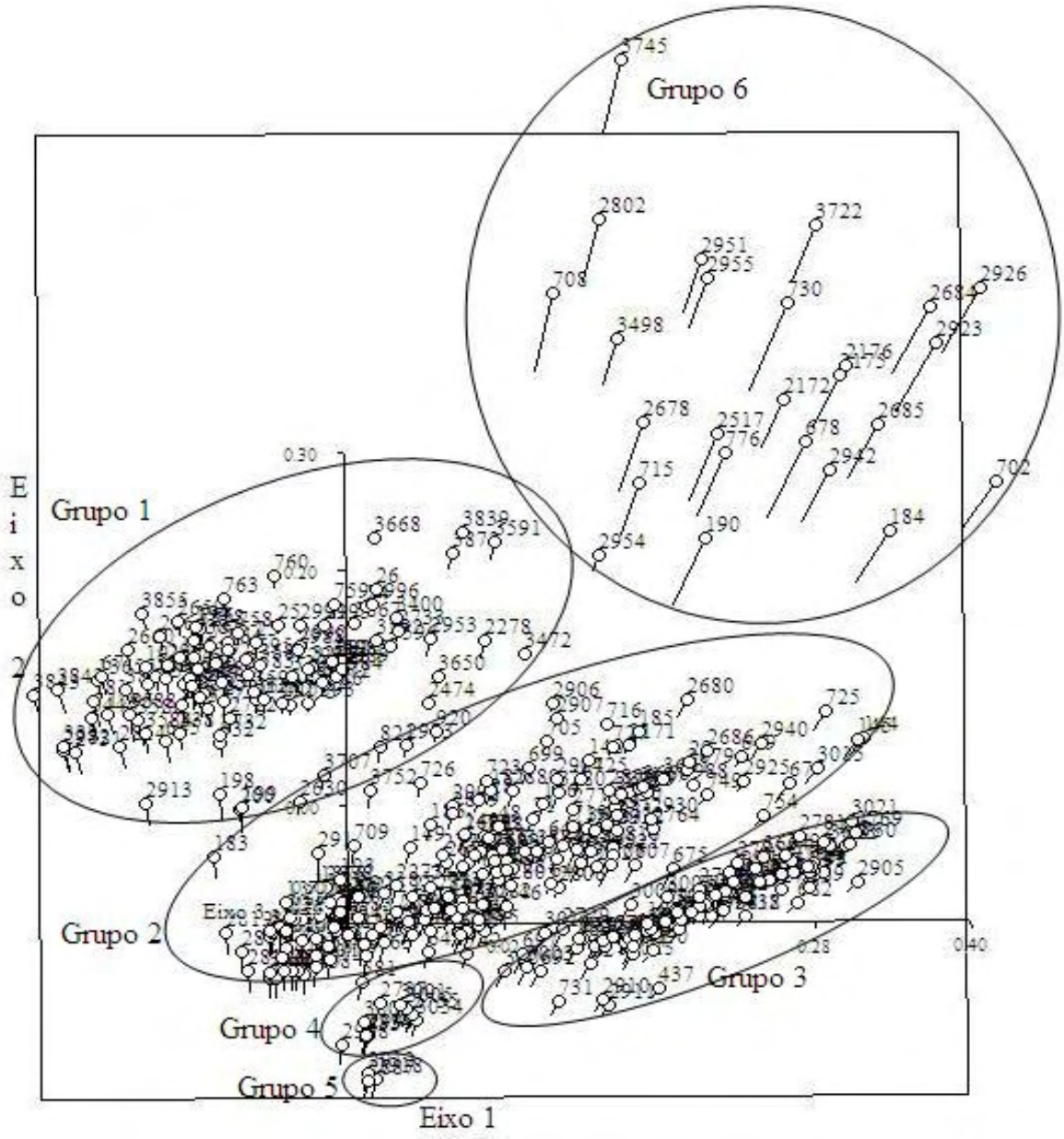

Figura 4.1. Dispersão de acessos da espécie *Capsicum annuum* nos três primeiros eixos fatoriais da análise de correspondência simples, aplicada à coleção de pimentas e pimentões da Embrapa Hortaliças.

Os acessos com maiores contribuições para o primeiro eixo fatorial da análise de correspondência foram: 43, 67, 144, 145, 181, 184, 189, 191, 195, 333, 682, 685, 701, 702, 703, 704, 724, 725, 1295, 1365, 2175, 2176, 2660, 2664, 2684, 2685, 2730, 2731, 2769, 2778, 2814, 2905, 2913, 2923, 2926, 2942, 2982, 3015, 3021, 3445, 3584, 3586, 3588, 3663, 3722, 3840, 3841, 3842, 3843 e 3855, representando mais de 39% da variabilidade do conjunto de dados nesse eixo (Tabela 4.2). Assim, é interessante que parte destes acessos estejam representados na coleção nuclear definida por este método. Os

indivíduos que mais contribuíram para o segundo eixo foram 24, 26, 184, 190, 433, 678, 702, 708, 715, 730, 731, 759, 760, 763, 776, 2172, 2175, 2176, 2517, 2678, 2684, 2685, 2734, 2735, 2737, 2802, 2887, 2912, 2918, 2923, 2926, 2942, 2951, 2954, 2955, 2996, 2998, 2999, 3001, 3002, 3005, 3034, 3498, 3591, 3668, 3722, 3745, 3839, 3867 e 3873, representando mais de 55% da variabilidade dos dados analisados neste eixo. E, os acessos com maiores contribuições para o terceiro eixo fatorial foram: 24, 184, 190, 291, 577, 582, 593, 594, 678, 702, 708, 715, 716, 730, 776, 988, 1375, 1378, 1379, 2172, 2175, 2176, 2474, 2517, 2678, 2684, 2685, 2737, 2802, 2887, 2912, 2918, 2922, 2923, 2926, 2942, 2951, 2955, 2996, 2998, 3001, 3002, 3472, 3596, 3650, 3668, 3722, 3745, 3839 e 3867, representando cerca de 55% da variabilidade neste eixo.

Tabela 4.2. Contribuição dos acessos da espécie *Capsicum annuum* para a variabilidade dos dados nos três primeiros eixos fatoriais da análise de correspondência, estimada a partir de 56 descritores qualitativos analisadas em 420 acessos da coleção de pimentas e pimentões da Embrapa Hortaliças.

Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3
23	0,27	0,26	0,03	704	0,62	0,03	0,20	2682	0,00	0,04	0,00	2959	0,52	0,19	0,10
24	0,06	0,44	0,68	705	0,03	0,04	0,16	2683	0,04	0,02	0,00	2982	0,94	0,02	0,02
25	0,21	0,34	0,32	706	0,10	0,05	0,09	2684	1,19	2,19	1,84	2983	0,02	0,04	0,19
26	0,05	0,49	0,39	707	0,14	0,05	0,02	2685	0,95	1,27	1,10	2984	0,11	0,19	0,41
29	0,25	0,02	0,36	708	0,03	2,07	2,69	2686	0,29	0,03	0,21	2985	0,10	0,23	0,39
30	0,00	0,11	0,02	709	0,08	0,02	0,04	2688	0,00	0,01	0,02	2986	0,15	0,15	0,29
31	0,02	0,00	0,13	710	0,27	0,09	0,01	2689	0,00	0,02	0,08	2987	0,16	0,15	0,31
32	0,07	0,26	0,03	712	0,01	0,12	0,04	2691	0,11	0,11	0,03	2988	0,17	0,24	0,19
33	0,09	0,11	0,07	713	0,00	0,09	0,04	2720	0,36	0,11	0,05	2989	0,10	0,21	0,40
40	0,34	0,06	0,02	714	0,00	0,10	0,16	2721	0,41	0,12	0,06	2990	0,17	0,26	0,20
41	0,39	0,04	0,10	715	0,16	0,93	1,28	2722	0,36	0,08	0,07	2991	0,15	0,31	0,36
42	0,36	0,08	0,11	716	0,11	0,07	0,47	2723	0,02	0,09	0,01	2992	0,22	0,12	0,24
43	0,66	0,04	0,31	717	0,57	0,04	0,18	2724	0,11	0,18	0,18	2993	0,20	0,11	0,29
44	0,54	0,04	0,12	718	0,42	0,13	0,10	2725	0,06	0,00	0,01	2994	0,16	0,20	0,23
45	0,53	0,05	0,12	719	0,06	0,00	0,15	2726	0,03	0,14	0,02	2995	0,12	0,32	0,41
47	0,00	0,02	0,05	720	0,09	0,20	0,00	2729	0,04	0,16	0,09	2996	0,05	0,42	0,57
48	0,00	0,01	0,00	721	0,23	0,13	0,01	2730	0,78	0,02	0,09	2997	0,35	0,05	0,09
64	0,17	0,00	0,39	722	0,12	0,00	0,06	2731	0,91	0,02	0,02	2998	0,10	0,61	0,54
66	0,44	0,21	0,10	723	0,00	0,00	0,04	2732	0,40	0,05	0,07	2999	0,07	0,57	0,33
67	0,82	0,16	0,01	724	0,64	0,04	0,05	2733	0,03	0,32	0,28	3001	0,03	0,41	0,45
68	0,37	0,16	0,14	725	0,71	0,07	0,00	2734	0,07	0,58	0,36	3002	0,07	0,51	0,50
69	0,48	0,16	0,11	726	0,02	0,00	0,05	2735	0,02	0,47	0,27	3003	0,01	0,27	0,08
143	0,24	0,14	0,06	727	0,11	0,03	0,26	2737	0,05	0,44	0,85	3005	0,03	0,42	0,32
144	0,87	0,04	0,00	728	0,00	0,00	0,06	2745	0,48	0,09	0,02	3006	0,00	0,28	0,03
145	0,80	0,03	0,00	729	0,00	0,18	0,00	2763	0,07	0,25	0,10	3007	0,19	0,09	0,01

Tabela 4.2. (Continuação)

Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3
146	0,02	0,00	0,03	730	0,50	1,91	3,91	2764	0,18	0,00	0,27	3009	0,12	0,11	0,01
147	0,12	0,13	0,13	731	0,04	0,38	0,03	2765	0,01	0,12	0,03	3011	0,02	0,19	0,27
148	0,08	0,02	0,26	732	0,16	0,16	0,20	2766	0,23	0,31	0,40	3013	0,16	0,16	0,06
149	0,02	0,02	0,00	733	0,00	0,01	0,00	2767	0,26	0,07	0,05	3015	0,65	0,01	0,01
150	0,10	0,11	0,19	734	0,00	0,09	0,01	2768	0,30	0,11	0,13	3018	0,33	0,07	0,08
180	0,00	0,13	0,10	735	0,00	0,18	0,01	2769	0,70	0,01	0,21	3019	0,37	0,06	0,15
181	0,85	0,09	0,08	736	0,00	0,13	0,00	2770	0,09	0,15	0,00	3020	0,08	0,19	0,02
183	0,37	0,04	0,15	737	0,00	0,17	0,01	2771	0,30	0,09	0,11	3021	0,71	0,00	0,08
184	1,00	0,61	1,72	738	0,00	0,15	0,01	2772	0,01	0,03	0,00	3024	0,12	0,18	0,05
185	0,16	0,05	0,20	739	0,01	0,05	0,04	2773	0,05	0,00	0,16	3027	0,22	0,13	0,08
186	0,01	0,01	0,00	741	0,00	0,15	0,02	2774	0,44	0,05	0,12	3029	0,22	0,14	0,11
187	0,00	0,00	0,06	742	0,00	0,23	0,00	2778	0,60	0,01	0,07	3031	0,31	0,10	0,13
188	0,16	0,15	0,01	743	0,01	0,09	0,19	2779	0,35	0,06	0,08	3032	0,23	0,12	0,09
189	0,64	0,07	0,09	744	0,01	0,17	0,03	2780	0,33	0,07	0,10	3033	0,21	0,10	0,05
190	0,26	0,45	3,60	746	0,01	0,12	0,02	2781	0,54	0,01	0,01	3034	0,02	0,50	0,13
191	0,62	0,18	0,00	747	0,00	0,17	0,01	2783	0,29	0,08	0,06	3037	0,09	0,20	0,01
192	0,38	0,20	0,13	748	0,00	0,05	0,02	2784	0,19	0,13	0,06	3038	0,22	0,15	0,07
193	0,10	0,07	0,24	749	0,29	0,00	0,14	2785	0,37	0,04	0,07	3183	0,02	0,02	0,10
194	0,41	0,29	0,08	750	0,00	0,11	0,00	2786	0,28	0,10	0,10	3184	0,02	0,03	0,01
195	0,61	0,14	0,05	751	0,11	0,19	0,05	2787	0,51	0,07	0,18	3185	0,05	0,09	0,00
196	0,48	0,22	0,12	752	0,23	0,33	0,32	2788	0,22	0,12	0,09	3186	0,00	0,02	0,07
198	0,37	0,00	0,18	753	0,12	0,22	0,25	2789	0,27	0,09	0,07	3187	0,15	0,00	0,32
199	0,31	0,00	0,20	754	0,44	0,00	0,16	2791	0,25	0,09	0,03	3189	0,14	0,00	0,36
200	0,32	0,00	0,16	755	0,40	0,30	0,15	2792	0,50	0,03	0,07	3280	0,04	0,00	0,01
291	0,13	0,04	0,73	756	0,39	0,25	0,17	2793	0,26	0,11	0,12	3281	0,00	0,04	0,01
292	0,08	0,10	0,06	757	0,34	0,27	0,20	2794	0,33	0,08	0,11	3282	0,00	0,01	0,00
295	0,16	0,14	0,18	758	0,31	0,30	0,23	2796	0,22	0,12	0,09	3294	0,26	0,12	0,11
296	0,20	0,23	0,28	759	0,11	0,42	0,21	2802	0,09	3,24	1,07	3295	0,19	0,18	0,25
297	0,21	0,17	0,18	760	0,20	0,49	0,26	2805	0,01	0,08	0,02	3296	0,15	0,14	0,18
332	0,40	0,04	0,09	761	0,43	0,18	0,07	2806	0,04	0,07	0,13	3297	0,20	0,18	0,24
333	0,94	0,03	0,03	763	0,34	0,39	0,02	2808	0,17	0,34	0,21	3298	0,23	0,19	0,25
337	0,41	0,09	0,02	764	0,34	0,25	0,18	2809	0,18	0,31	0,27	3299	0,15	0,20	0,27
338	0,52	0,05	0,04	775	0,02	0,18	0,08	2810	0,26	0,29	0,42	3377	0,14	0,17	0,32
432	0,43	0,13	0,12	776	0,31	0,90	2,17	2811	0,01	0,14	0,01	3400	0,03	0,38	0,41
433	0,07	0,58	0,33	777	0,03	0,12	0,06	2812	0,14	0,29	0,33	3445	0,79	0,06	0,01
437	0,18	0,32	0,03	778	0,05	0,23	0,15	2813	0,19	0,25	0,41	3472	0,02	0,24	0,63
577	0,11	0,16	0,62	821	0,05	0,03	0,04	2814	0,71	0,02	0,03	3498	0,10	1,77	0,36
580	0,07	0,24	0,40	822	0,18	0,15	0,26	2815	0,02	0,13	0,02	3581	0,26	0,19	0,04
581	0,08	0,36	0,21	823	0,00	0,10	0,01	2816	0,10	0,02	0,17	3582	0,27	0,11	0,15
582	0,10	0,21	0,54	920	0,01	0,05	0,17	2817	0,13	0,15	0,21	3583	0,28	0,17	0,16
583	0,08	0,16	0,33	987	0,00	0,03	0,01	2818	0,30	0,24	0,25	3584	0,63	0,04	0,00

Tabela 4.2. (Continuação)

Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3
593	0,17	0,22	0,47	988	0,29	0,01	0,43	2819	0,32	0,18	0,32	3585	0,43	0,11	0,00
594	0,15	0,22	0,47	990	0,06	0,05	0,07	2827	0,18	0,31	0,26	3586	0,68	0,06	0,16
602	0,11	0,12	0,02	991	0,11	0,00	0,03	2829	0,10	0,00	0,22	3587	0,47	0,13	0,03
640	0,03	0,16	0,26	994	0,03	0,16	0,01	2830	0,11	0,00	0,32	3588	0,71	0,06	0,21
641	0,01	0,24	0,04	1295	0,79	0,07	0,01	2831	0,10	0,00	0,29	3589	0,45	0,15	0,00
642	0,00	0,19	0,03	1358	0,44	0,32	0,23	2878	0,12	0,08	0,01	3590	0,46	0,14	0,04
643	0,00	0,05	0,00	1365	0,83	0,15	0,02	2879	0,00	0,00	0,02	3591	0,00	0,75	0,09
644	0,03	0,02	0,01	1375	0,15	0,11	0,47	2880	0,01	0,00	0,01	3592	0,05	0,30	0,41
645	0,53	0,03	0,01	1376	0,12	0,12	0,28	2887	0,06	0,77	0,57	3593	0,16	0,12	0,40
646	0,03	0,06	0,11	1377	0,18	0,16	0,23	2891	0,22	0,33	0,35	3594	0,28	0,13	0,10
649	0,43	0,02	0,19	1378	0,14	0,10	0,59	2892	0,06	0,18	0,03	3595	0,29	0,20	0,05
674	0,13	0,22	0,09	1379	0,20	0,12	0,50	2893	0,20	0,31	0,30	3596	0,04	0,28	0,68
675	0,20	0,05	0,07	1387	0,41	0,22	0,04	2896	0,11	0,26	0,10	3597	0,11	0,22	0,37
676	0,10	0,19	0,05	1401	0,00	0,04	0,01	2903	0,00	0,00	0,02	3650	0,01	0,19	0,60
677	0,56	0,00	0,03	1404	0,00	0,01	0,00	2904	0,04	0,01	0,26	3661	0,21	0,01	0,42
678	0,53	0,89	3,66	1405	0,07	0,02	0,08	2905	0,79	0,06	0,08	3662	0,11	0,00	0,29
680	0,00	0,04	0,01	1716	0,05	0,16	0,08	2906	0,04	0,10	0,31	3663	0,67	0,02	0,18
682	0,57	0,10	0,10	2171	0,14	0,04	0,32	2907	0,04	0,08	0,32	3664	0,43	0,03	0,06
683	0,25	0,12	0,00	2172	0,54	1,45	0,58	2910	0,08	0,36	0,00	3665	0,07	0,01	0,18
684	0,04	0,03	0,01	2175	0,79	1,70	0,82	2911	0,09	0,38	0,00	3666	0,16	0,00	0,29
685	0,74	0,02	0,27	2176	0,81	1,76	0,95	2912	0,06	0,72	0,54	3668	0,05	0,76	0,89
686	0,48	0,06	0,14	2272	0,10	0,10	0,10	2913	0,60	0,00	0,09	3707	0,13	0,01	0,05
687	0,12	0,19	0,00	2273	0,03	0,08	0,01	2918	0,06	0,78	0,44	3722	0,70	3,20	0,70
688	0,12	0,17	0,16	2275	0,55	0,16	0,08	2922	0,10	0,23	0,44	3745	0,12	5,07	1,27
689	0,06	0,19	0,00	2277	0,47	0,14	0,04	2923	1,20	1,82	2,10	3752	0,06	0,00	0,00
690	0,18	0,23	0,00	2278	0,00	0,23	0,05	2925	0,39	0,01	0,21	3753	0,52	0,03	0,26
691	0,43	0,06	0,10	2282	0,00	0,11	0,01	2926	1,28	2,07	1,24	3831	0,25	0,21	0,21
692	0,02	0,28	0,07	2283	0,14	0,26	0,19	2927	0,08	0,04	0,07	3832	0,23	0,20	0,22
693	0,05	0,17	0,04	2474	0,01	0,13	0,57	2928	0,00	0,06	0,00	3834	0,01	0,05	0,02
694	0,12	0,22	0,03	2517	0,31	1,07	2,45	2930	0,16	0,00	0,13	3836	0,01	0,03	0,14
695	0,33	0,13	0,04	2519	0,02	0,16	0,04	2931	0,05	0,02	0,07	3839	0,00	0,85	0,86
696	0,16	0,15	0,05	2521	0,10	0,02	0,15	2932	0,11	0,01	0,12	3840	0,75	0,15	0,10
697	0,01	0,24	0,13	2630	0,17	0,00	0,02	2933	0,07	0,01	0,01	3841	0,62	0,18	0,11
698	0,05	0,20	0,03	2659	0,50	0,34	0,15	2940	0,41	0,03	0,04	3842	1,03	0,12	0,02
699	0,02	0,01	0,01	2660	0,69	0,24	0,07	2942	0,64	0,85	0,96	3843	1,15	0,12	0,00
700	0,17	0,17	0,00	2664	0,57	0,29	0,12	2951	0,27	2,43	0,55	3855	0,60	0,36	0,12
701	0,64	0,04	0,23	2678	0,15	1,13	2,83	2953	0,01	0,23	0,03	3867	0,08	0,41	1,04
702	1,45	0,82	0,81	2679	0,25	0,01	0,42	2954	0,09	0,57	0,13	3872	0,12	0,16	0,41
703	0,61	0,04	0,19	2680	0,22	0,08	0,02	2955	0,29	2,36	0,46	3873	0,00	0,68	0,13

A partir desses resultados, amostrou-se aleatoriamente dentro de cada grupo identificado e de forma proporcional ao tamanho de cada grupo. Considerando-se a intensidade de amostragem estabelecida (10%), dos 420 acessos da coleção de *C. annuum*, sortearam-se 42, na seguinte proporção: doze acessos no grupo 1, quatorze no grupo 2, nove no grupo 3, um no grupo 4, um no grupo 5 e cinco no grupo 6. Desse modo, a coleção nuclear de *C. annuum*, segundo a abordagem da análise de correspondência, ficou assim constituída: 40, 149, 184, 186, 187, 291, 292, 338, 437, 641, 688, 702, 715, 716, 725, 730, 756, 778, 988, 1387, 2273, 2720, 2729, 2737, 2773, 2785, 2805, 2810, 2887, 2905, 2913, 2990, 2992, 3006, 3011, 3020, 3034, 3583, 3596, 3668, 3745 e 3843.

Pela análise de agrupamento do vizinho mais próximo, associada à matriz de distâncias entre os acessos, para variáveis multicategóricas, pôde-se verificar que os acessos dessa espécie são relativamente similares, haja vista a maior parte deles ter se agrupado em distâncias inferiores a 0,20 (Figura 4.2). Verificou-se também que os últimos acessos a entrarem na formação do dendrograma correspondem, em sua maioria, aos mais afastados das nuvens de pontos mais compactas observadas na Figura 4.1. Isso mostra certa concordância entre os resultados da análise de correspondência e os obtidos por esse método de agrupamento.

Os acessos 333 e 2912 mostraram tamanha similaridade por este método (Figura 4.2), que, possivelmente, tratam-se de duplicatas. A confirmação da duplicação de acessos, entretanto, não é passível de ser atestada a partir de dados fenotípicos, como ocorreu neste caso. Logo, isso requer análises específicas, como, por exemplo, a genotipagem dos acessos via marcadores moleculares, os quais revelam o polimorfismo genético em nível de DNA, livre de interferências ambientais (Ferreira & Grattapaglia, 1996). Isso reforça, também, a importância da caracterização molecular das coleções de germoplasma.

O corte do dendrograma para a constituição dos estratos a serem amostrados foi realizado na distância de ligação 0,22, o que resultou em 42 grupos, sendo o maior deles com 371 acessos, um grupo com cinco acessos, quatro grupos com dois acessos e 36 grupos de acessos isolados. Independentemente do tamanho dos estratos, como o número deles correspondeu exatamente ao tamanho da coleção nuclear a ser constituída (10% dos 420 acessos de *C. annuum*), tomou-se um acesso dentro de cada grupo, com o cuidado de sorteá-lo nos grupos com dois ou mais acessos. Assim, os acessos selecionados para representar a coleção de *C. annuum*, segundo este método, foram: 184, 188, 674, 678, 685,

695, 699, 706, 715, 724, 733, 734, 749, 990, 991, 1376, 2175, 2176, 2277, 2278, 2683, 2684, 2688, 2725, 2769, 2785, 2802, 2817, 2905, 2923, 2926, 2942, 2953, 2955, 2988, 3400, 3498, 3595, 3707, 3722, 3745 e 3836. Destes, 14% foram também selecionados para compor a coleção estabelecida pela análise de correspondência, o que corresponde a seis acessos coincidentes nas duas coleções.

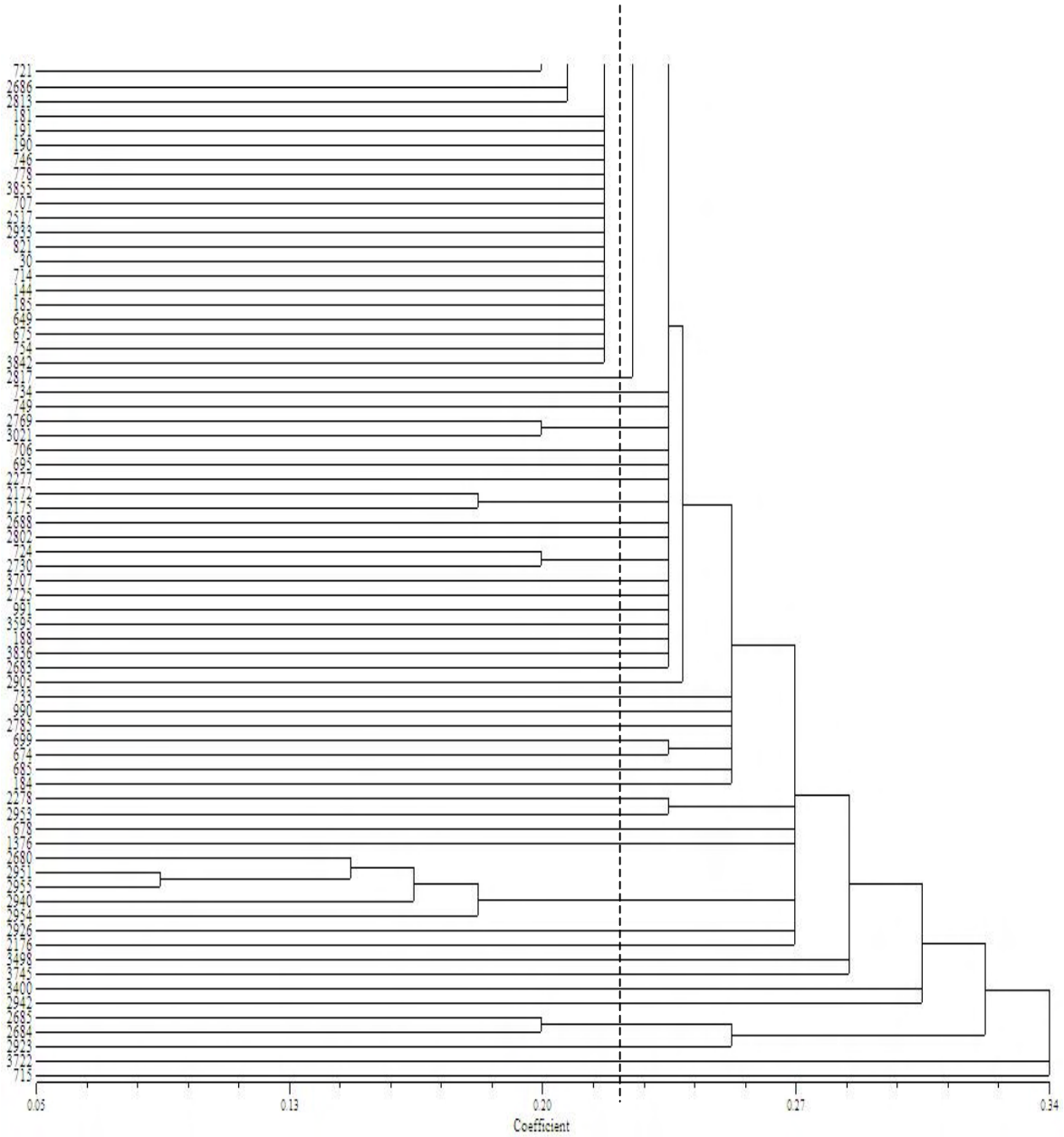

Figura 4.2. Dendrograma ilustrativo do padrão de dissimilaridade entre 420 acessos de *Capsicum annuum* da coleção de pimentas e pimentões da Embrapa Hortaliças estimado pelo método hierárquico do vizinho mais próximo com base na distância para variáveis multicategóricas estimada a partir de 56 descritores qualitativos (continua).

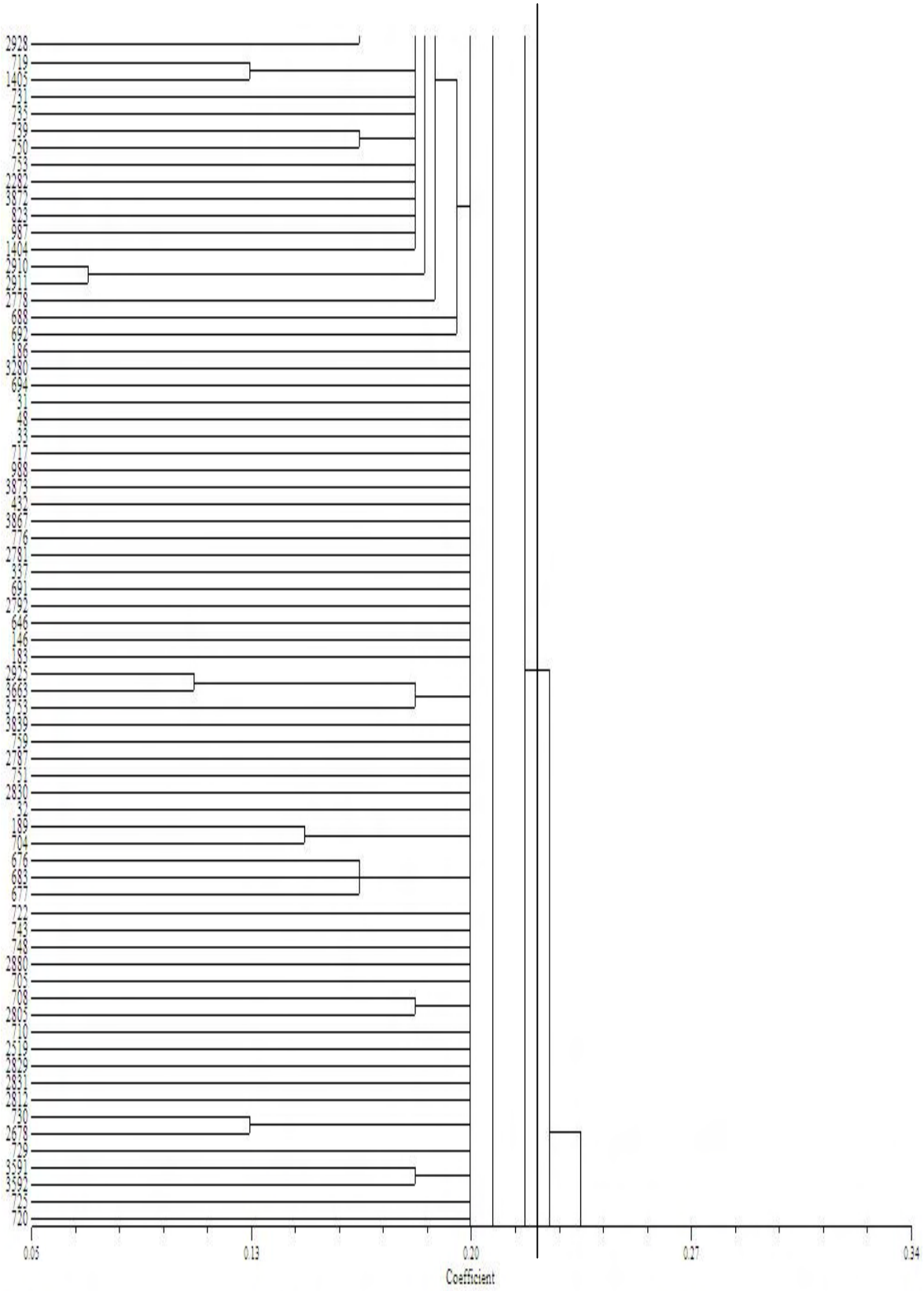

Figura 4.2. (Continuação)

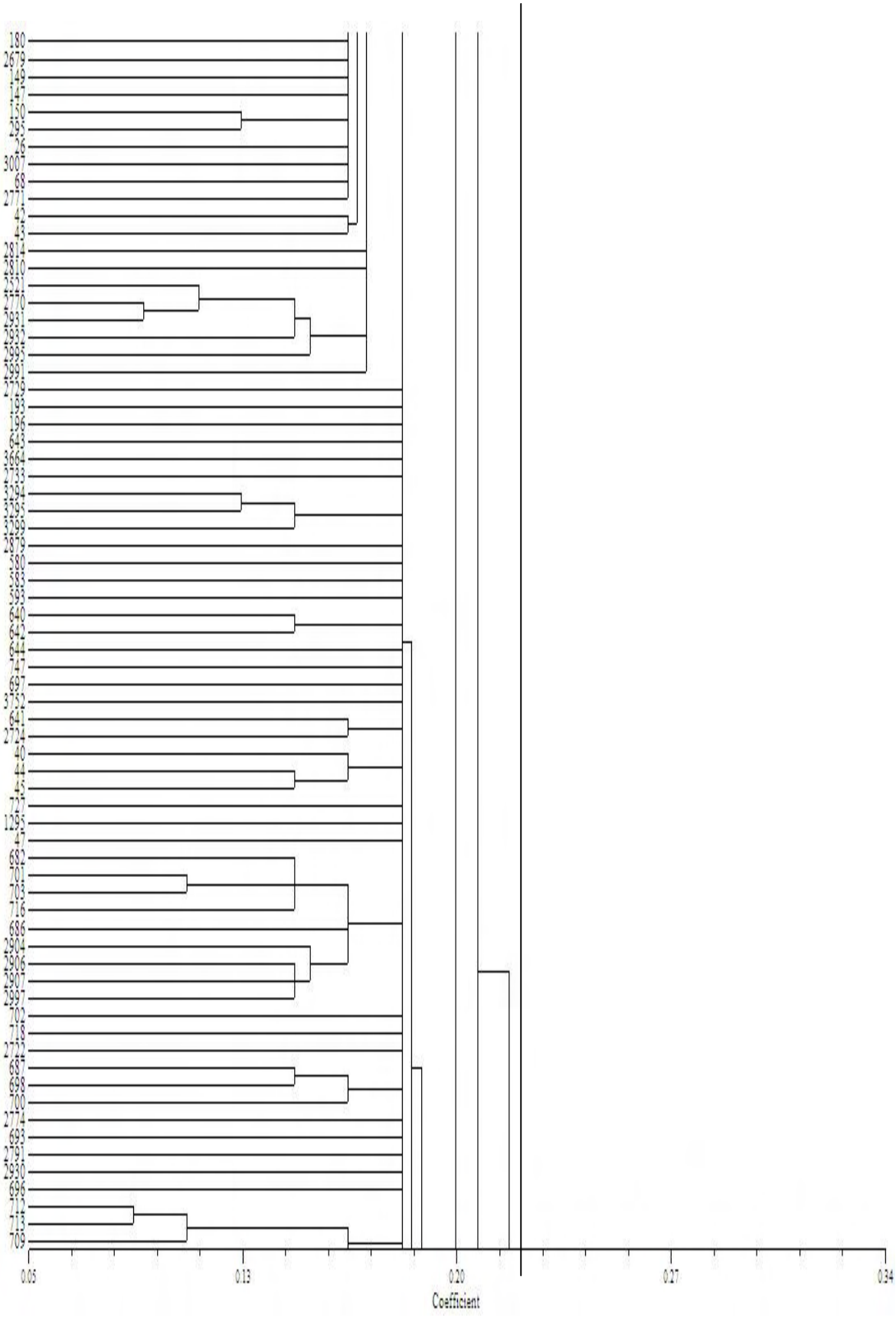

Figura 4.2. (Continuação)

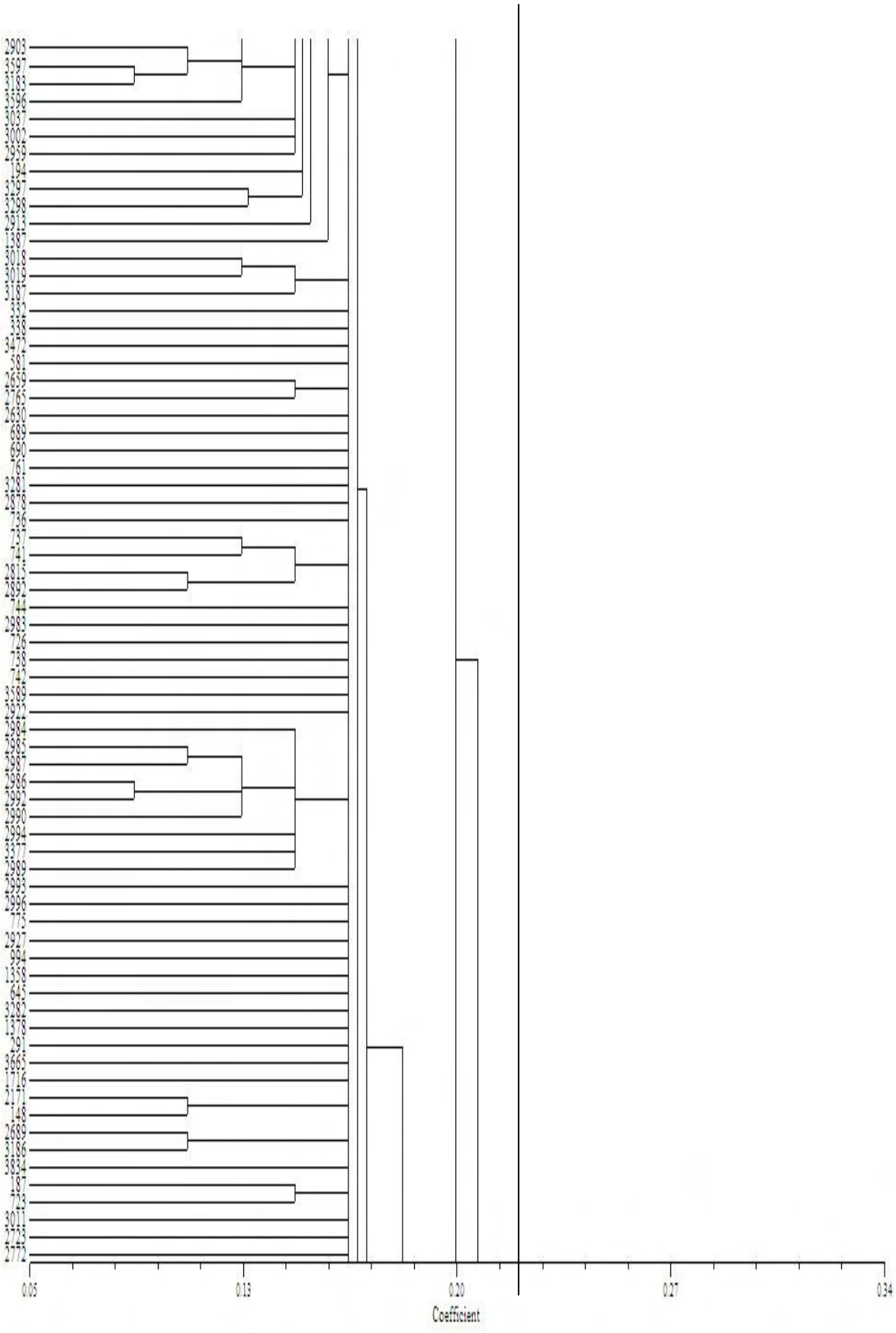

Figura 4.2. (Continuação)

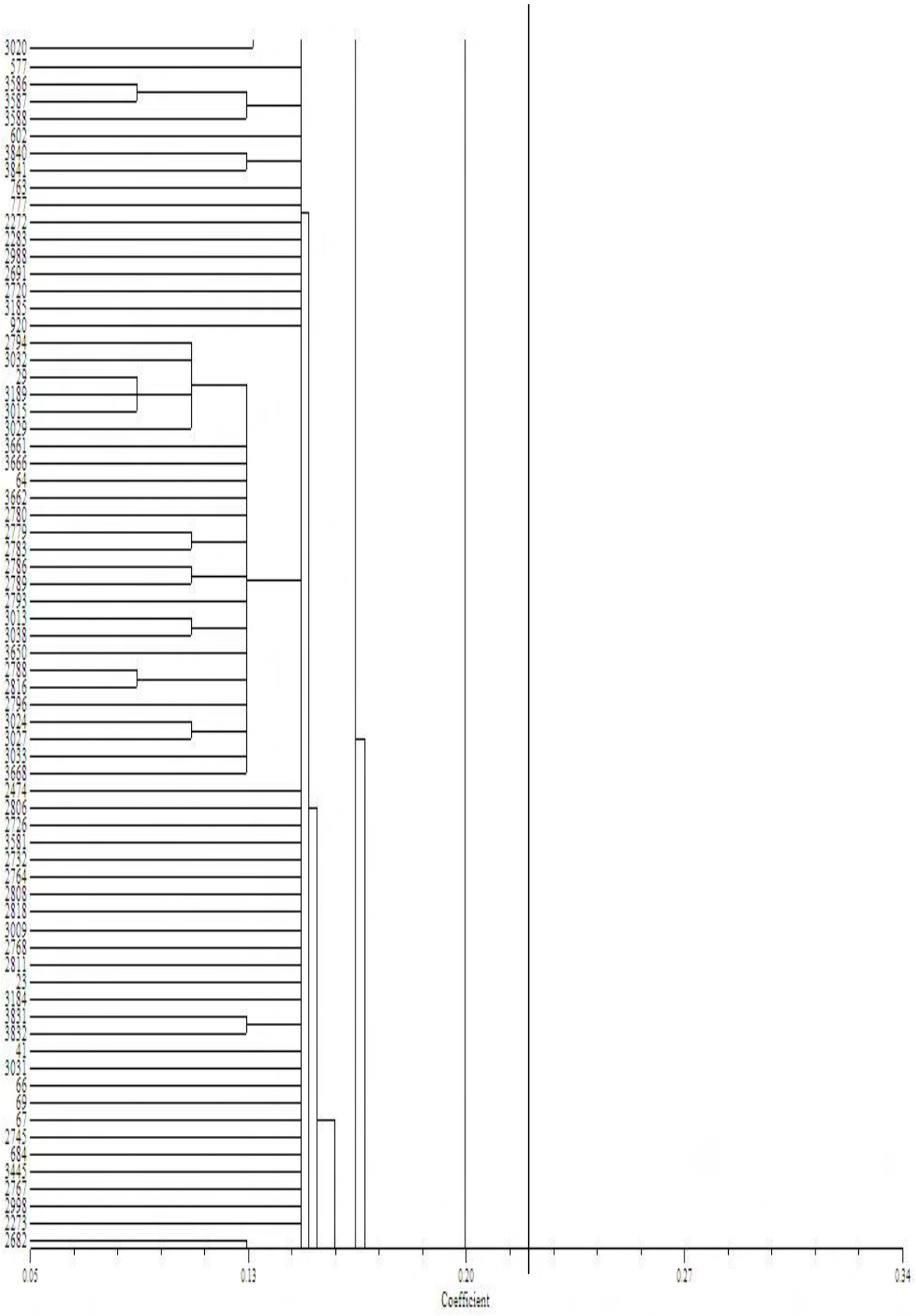

Figura 4.2. (Continuação)

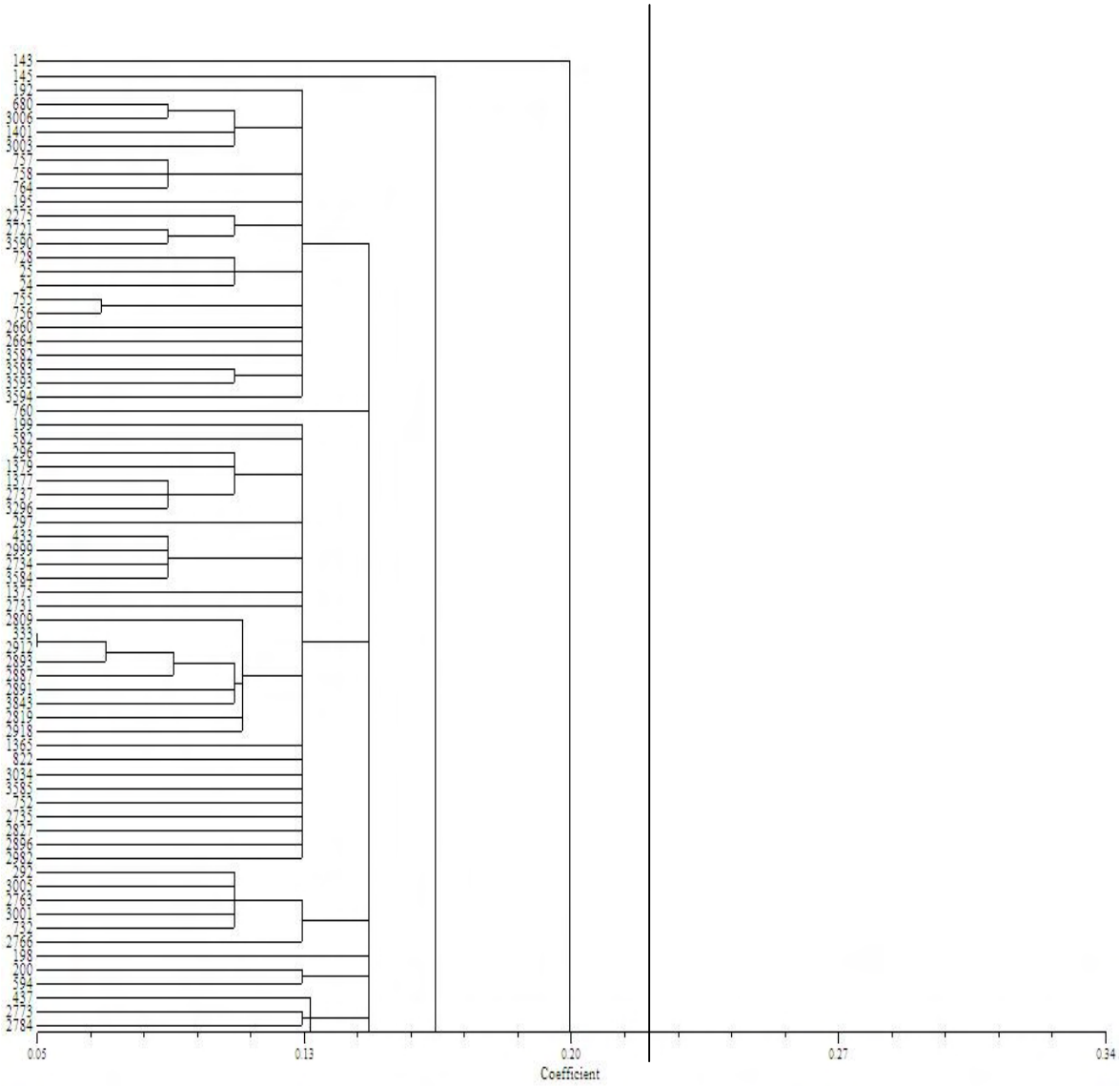

Figura 4.2. (Continuação)

A partir da utilização do método de Tocher, com critério de aglomeração inverso, considerando-se a intensidade de amostragem de 10%, os seguintes acessos foram selecionados para compor a coleção nuclear de *C. annuum*: 32, 44, 191, 192, 196, 291, 292, 433, 644, 674, 725, 741, 823, 1365, 1716, 2683, 2691, 2720, 2729, 2737, 2763, 2766, 2783, 2806, 2810, 2813, 2829, 2878, 2912, 2995, 2996, 3001, 3002, 3033, 3472, 3587, 3589, 3590, 3591, 3593, 3650 e 3707. Esta coleção teve concordância de 17% com a coleção nuclear obtida por análise de correspondência e de 7% com a coleção nuclear obtida pelo método de agrupamento do vizinho mais próximo.

Após a obtenção das coleções nucleares, a partir dos três métodos, fez-se a validação dessas coleções segundo os critérios estabelecidos (CR% e DF). Verificou-se,

então, que o coeficiente de similaridade (CR%) entre a coleção nuclear obtida pela análise de correspondência e a coleção inicial foi de 79,3%; utilizando-se a coleção nuclear obtida pelo agrupamento do vizinho mais próximo essa similaridade foi de 89,2%; e, por último, a coleção nuclear resultante do método de Tocher, com critério de aglomeração inverso, teve 76,5% de similaridade com a coleção inicial. Segundo Hu et al. (2000), para que a coleção nuclear seja considerada representativa da coleção inicial, deve-se buscar o limite mínimo de 80% para esse coeficiente de similaridade. E, neste caso, apenas o método de agrupamento pelo vizinho mais próximo permitiu este valor.

De acordo com o índice de diversidade fenotípica (DF), as coleções nucleares sugeridas pela análise de correspondência, pelo método de agrupamento do vizinho mais próximo e pelo método de Tocher resultaram nos seguintes coeficientes: 1,04, 1,28 e 0,99, respectivamente. A ordem destes coeficientes correspondeu, exatamente, à ordem daqueles obtidos segundo o primeiro critério (CR%), isto é, a melhor proposta de coleção nuclear é a do método do vizinho mais próximo, seguida da análise de correspondência e, por último, do método de Tocher invertido. A vantagem do método do vizinho mais próximo, neste caso, foi relativamente alta, sobretudo em relação ao método de Tocher. Primeiramente porque foi o único com mais de 80% de similaridade entre a sua coleção nuclear e a coleção base, e, em segundo lugar, porque sua coleção nuclear revela, nitidamente, distribuição mais equilibrada dos acessos nas categorias das variáveis, relativamente à coleção base (DF = 1,28). As razões dessa superioridade relativa, bem como do pior desempenho do método de Tocher, não foram aqui identificados, necessitando de estudos específicos, envolvendo possivelmente simulações e, ou, métodos computacionalmente intensivos.

Segundo van Hintum et al. (2000), uma expectativa no estabelecimento de uma coleção nuclear é que as variâncias aumentem nesta coleção comparativamente à coleção base. De acordo com Oliveira (2007), o aumento da variância na coleção nuclear em relação à coleção base é esperado, uma vez que nas coleções nucleares são selecionados acessos mais divergentes, ou seja, a coleção nuclear é selecionada para conter uma maior proporção dos tipos extremos do que dos tipos comuns, resultando em maior variância na coleção nuclear do que na coleção base. Esse aumento na variância e redução no número de acessos na coleção nuclear, provavelmente, significam manutenção da variabilidade genética da coleção base. Nesse aspecto, os outros dois métodos, com índices DF em torno de 1,0, revelaram distribuição similar à coleção inicial. Por esta razão, a proposta final de

coleção nuclear para o germoplasma de *C. annuum* ficou constituída pelos seguintes acessos: 184, 188, 674, 678, 685, 695, 699, 706, 715, 724, 733, 734, 749, 990, 991, 1376, 2175, 2176, 2277, 2278, 2683, 2684, 2688, 2725, 2769, 2785, 2802, 2817, 2905, 2923, 2926, 2942, 2953, 2955, 2988, 3400, 3498, 3595, 3707, 3722, 3745 e 3836.

4.3.2 Coleção nuclear para *Capsicum baccatum*

A dispersão gráfica da coleção de acessos de *C. baccatum* (106 acessos) nos três primeiros eixos fatoriais da análise de correspondência (Figura 4.3) revela a formação de um grupo que concentra a maior parte dos acessos dessa espécie (grupo 1). Esses eixos fatoriais explicaram 47% da variabilidade do conjunto de dados. Os acessos 2978, 2948, 2964, 3699, 3825, 3817, 3811, 2956, 3750, 2935, 639 e 288 são os mais dispersos entre si e do grupo central, indicando apresentarem características peculiares e divergentes da maior parte dos acessos. Esses acessos merecem, portanto, fazer parte da coleção nuclear a ser constituída segundo este método e, por isso, foram distribuídos em outros sete grupos menores, a maioria com um a três acessos (Figura 4.3), para garantir sua representação na amostra de acessos a ser selecionada.

Os acessos 2948 e 3838, em relação ao primeiro eixo fatorial, e 3750 e 288, em relação ao segundo, formam os contrastes de acessos mais divergentes identificados nesta análise. O primeiro par se diverge em 32 descritores, com destaque para comprimento do fruto, com até 1 cm para 2948 contra 4-8 cm de 3838; espessura da parede do fruto de até 1 mm para 2948, e de 4-5 mm para 3838; largura do fruto de até 1 cm para 2948 e de 5-8 cm em 3838; e peso do fruto de até 1 g para 2948 e de 27-81 g para 3838; e pungência em que 2948 possui fruto altamente picante e 3838 possui fruto doce. Os acessos 3750 e 288 diferem em 23 descritores, com destaque para o formato do fruto, em que o acesso 288 possui frutos alongados e o acesso 3750 possui frutos tipo pitanga. Isso influencia no peso do fruto, sendo que o primeiro acesso apresenta frutos com peso de 1-3 g e o segundo de 9-27 g. Com relação à procedência o primeiro vem da região Centro-Oeste do Brasil (288) e o outro da região Sudeste (3750). Isso mostra que os caracteres do fruto são determinantes na composição dos dois primeiros eixos fatoriais da análise de correspondência e, por conseguinte, são também muito importantes para a discriminação dos acessos dessa espécie (Reifschneider, 2000).

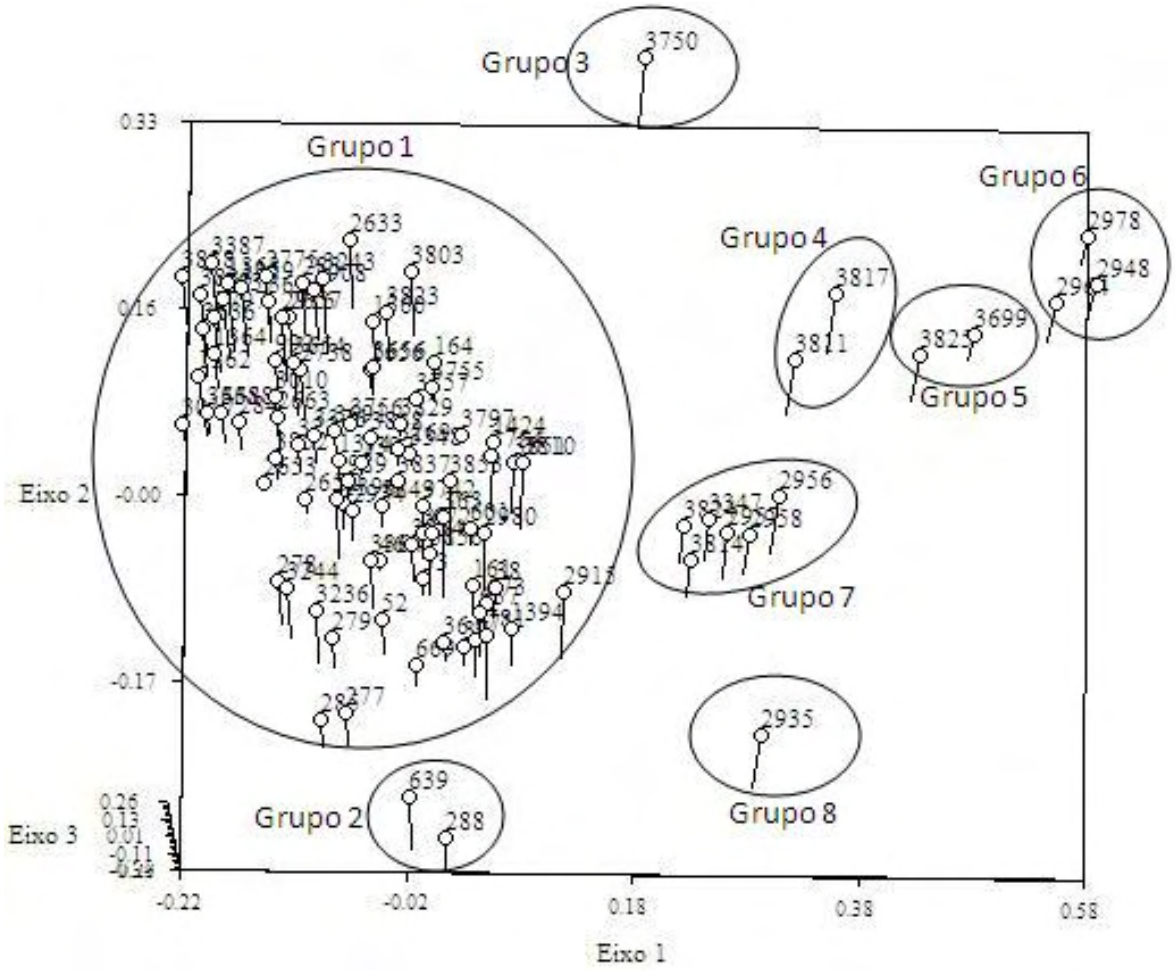

Figura 4.3. Dispersão de acessos da espécie *Capsicum baccatum* nos três primeiros eixos fatoriais da análise de correspondência simples, aplicada à coleção de pimentas e pimentões da Embrapa Hortaliças.

Os acessos com maiores contribuições para o primeiro eixo fatorial da análise de correspondência foram: 1362, 2935, 2948, 2956, 2957, 2958, 2964, 2978, 3017, 3347, 3636, 3638, 3699, 3750, 3811, 3814, 3817, 3824, 3825, 3833 e 3838, representando 78% da variabilidade captada neste eixo (Tabela 4.3). Os que mais contribuíram para o segundo eixo fatorial foram: 36, 37, 277, 279, 283, 285, 288, 637, 639, 669, 671, 781, 1394, 2633, 2915, 2935, 2948, 2964, 2978, 3236 e 3750, representando mais de 60% da variabilidade associada a esse eixo. E, no terceiro eixo, os acessos 46, 363, 781, 1359, 1397, 1424, 2653, 2655, 2915, 3243, 3348, 3754, 3755, 3756, 3757, 3797, 3810, 3829, 3837, 3848 e 3853 apresentaram as maiores contribuições. Para a composição da coleção nuclear, os acessos selecionados devem representar bem a variabilidade da coleção base, logo, é interessante que alguns destes acessos sejam amostrados. Assim, para garantir boa representação da coleção inicial, foi realizada amostragem aleatória estratificada proporcional dentro de

cada um dos oito grupos identificados. E, os acessos selecionados neste método, para composição da coleção nuclear de *C. baccatum*, foram: 37, 52, 53, 285, 288, 362, 363, 781, 1360, 1362, 1364, 2633, 2653, 2935, 2948, 2958, 2980, 3236, 3347, 3656, 3699, 3750, 3797, 3803, 3817, 3833, 3851 e 3856.

Tabela 4.3. Contribuição dos acessos da espécie *Capsicum baccatum* para a variabilidade dos dados nos três primeiros eixos fatoriais da análise de correspondência, estimada a partir de 56 descritores qualitativos analisadas em 106 acessos da coleção de pimentas e pimentões da Embrapa Hortaliças.

Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3
36	0,01	1,57	1,04	989	0,14	0,02	0,11	2956	2,79	0,17	1,05	3750	1,21	8,51	1,36
37	0,03	1,62	1,11	992	0,50	0,24	0,74	2957	2,04	0,35	0,19	3754	0,09	0,06	2,78
38	0,11	0,81	0,14	1359	0,84	0,99	1,60	2958	2,43	0,31	0,00	3755	0,00	0,48	2,02
39	0,00	0,15	1,57	1360	0,08	0,90	0,00	2964	10,62	1,54	0,03	3756	0,15	0,19	2,50
46	0,06	0,32	2,21	1362	1,27	0,17	0,85	2978	11,35	3,31	0,66	3757	0,01	0,48	4,04
50	0,21	0,25	1,47	1363	0,96	1,06	1,48	2980	0,07	0,48	1,32	3797	0,02	0,18	3,75
52	0,05	1,32	0,00	1364	1,12	0,58	0,10	3010	0,49	0,08	0,32	3803	0,01	1,41	1,00
53	0,00	0,58	1,13	1374	0,19	0,00	0,00	3017	1,29	0,09	0,60	3810	0,20	0,06	1,94
161	0,05	0,71	0,49	1394	0,18	1,69	0,04	3169	0,02	0,04	1,39	3811	2,91	0,34	0,49
164	0,00	0,44	0,01	1396	0,16	0,06	0,40	3236	0,28	1,50	0,84	3814	1,63	0,53	0,00
277	0,16	3,40	0,01	1397	0,11	0,02	2,16	3237	0,38	0,02	0,27	3817	3,93	1,17	0,57
278	0,53	0,95	0,15	1424	0,10	0,02	2,61	3243	0,27	1,02	2,45	3823	0,04	0,92	0,07
279	0,20	1,56	0,08	2633	0,16	1,92	1,55	3244	0,44	1,06	0,61	3824	1,57	0,24	0,00
283	0,41	1,42	0,13	2653	0,69	0,00	3,31	3347	1,81	0,22	0,04	3825	6,16	0,55	0,08
284	0,89	0,10	0,40	2655	0,40	0,03	1,62	3348	0,01	0,04	2,13	3829	0,02	0,17	1,74
285	0,26	3,52	0,09	2656	0,09	0,59	1,30	3364	0,08	0,76	0,50	3833	1,25	1,03	0,66
288	0,01	7,64	0,09	2657	0,49	0,84	0,27	3376	0,31	0,04	0,36	3837	0,02	0,00	2,15
362	0,65	1,38	0,27	2663	0,57	0,03	0,33	3387	1,12	1,39	1,45	3838	1,51	1,34	0,71
363	0,01	0,62	5,44	2738	0,39	0,28	0,18	3636	1,25	0,72	0,10	3848	0,09	0,11	2,23
601	0,04	0,13	1,20	2775	0,63	1,16	1,45	3637	1,05	0,99	0,71	3849	0,05	0,05	0,89
637	0,06	1,43	0,17	2797	1,12	0,78	0,36	3638	1,20	0,17	0,85	3851	0,18	0,02	0,38
639	0,01	6,36	1,50	2908	0,34	1,30	0,21	3654	0,41	0,30	0,38	3852	0,54	0,01	0,69
669	0,00	1,96	0,73	2915	0,46	1,50	2,65	3656	0,08	0,48	0,28	3853	0,01	0,13	1,61
671	0,06	1,84	0,02	2916	0,47	0,61	1,05	3658	1,00	0,12	0,22	3854	0,01	0,45	0,00
672	0,00	0,16	1,52	2934	0,13	0,11	0,20	3691	0,20	0,05	0,16	3856	0,00	0,62	0,05
673	0,08	0,95	0,66	2935	2,41	4,19	1,23	3699	7,48	1,10	0,59	–	–	–	–
781	0,08	2,17	2,82	2948	11,96	1,95	0,06	3742	0,00	0,21	0,48	–	–	–	–

Utilizando-se o método de agrupamento do vizinho mais próximo e um corte no dendrograma obtido, na distância de ligação de 0,24 (Figura 4.4), obteve-se a formação de um grupo maior, com setenta acessos, um grupo com doze, um grupo com cinco, um grupo com quatro, dois grupos com dois acessos e onze de acessos isolados.

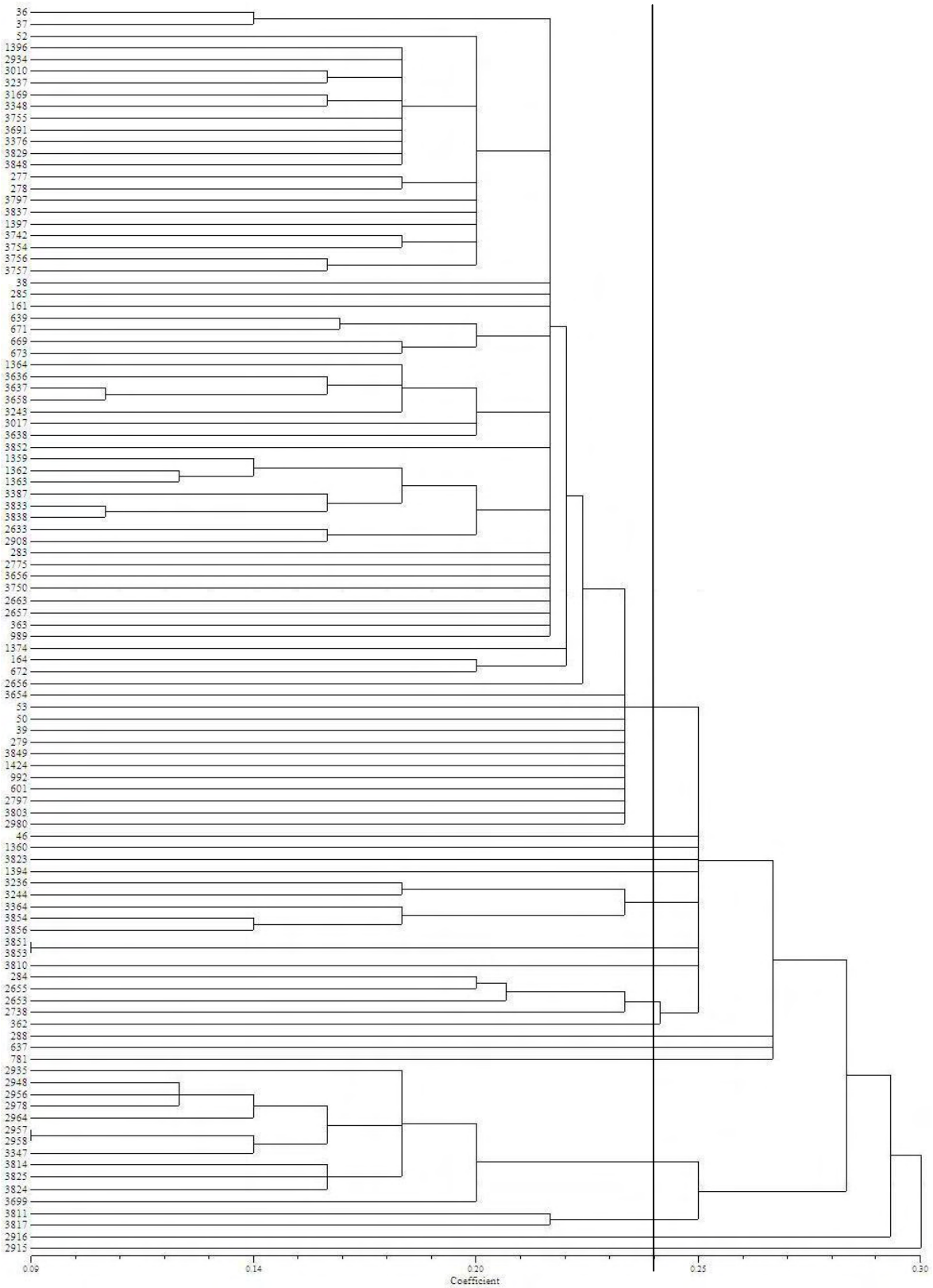

Figura 4.4. Dendrograma ilustrativo do padrão de dissimilaridade pelo método hierárquico do vizinho mais próximo com base na distância para variáveis multicategóricas estimada a partir de 56 descritores qualitativos analisadas em 106 acessos de *Capsicum baccatum* da coleção de pimentas e pimentões da Embrapa Hortaliças.

Devido à alta similaridade observada entre os acessos 2957 e 2958, bem como entre 3851 e 3853, há forte indício que estes pares de acessos tratem-se de possíveis duplicatas. Como mencionado anteriormente, a confirmação dessas duplicações, contudo, deve passar por análise específica, incluindo, por exemplo, a genotipagem por marcadores moleculares.

Para constituir a coleção nuclear de *C. baccatum* por este método, dentro de cada grupo com dois ou mais acessos, realizaram-se amostragens aleatórias proporcionais, considerando-se a intensidade estabelecida de 25% e garantindo-se a representação dos estratos menores com, pelo menos, um acesso. Assim, selecionaram-se 28 acessos, sendo um por grupo de até cinco acessos (quinze grupos), dois para o grupo de doze acessos e onze para o grupo maior, resultando na seguinte relação: 36, 46, 52, 53, 277, 288, 362, 601, 637, 781, 1360, 1394, 2655, 2656, 2657, 2775, 2915, 2916, 2956, 3017, 3364, 3755, 3810, 3811, 3823, 3825, 3837 e 3851. Destes, 21% foram também selecionados para compor a coleção estabelecida pela análise de correspondência, o que equivale a seis acessos em comum nas duas coleções.

Pela utilização do método de Tocher, com critério de aglomeração inverso, considerando-se a intensidade de amostragem de 25%, os seguintes acessos foram selecionados para a composição da coleção nuclear e *C. baccatum*: 283, 637, 673, 989, 992, 1363, 1397, 2633, 2653, 2775, 2958, 2964, 2978, 3169, 3243, 3347, 3636, 3691, 3699, 3742, 3797, 3803, 3810, 3811, 3814, 3837, 3853 e 3856. Esta coleção teve concordância de 21% com a coleção obtida por análise de correspondência e de 18% com a coleção nuclear obtida pelo método de agrupamento do vizinho mais próximo.

Os coeficientes de similaridade (CR%) entre as coleções nucleares, obtidas pela análise de correspondência, pelo agrupamento do vizinho mais próximo e pelo método de Tocher, e a coleção inicial foram 92,1%, 91,5% e 88,5%, respectivamente. Logo, neste caso, todas as coleções propostas atingiram o valor mínimo de similaridade (80%) preconizado por Hu et al. (2000), para que uma coleção nuclear seja representativa da coleção inicial. A vantagem, embora pequena, neste caso recaiu sobre a análise de correspondências, e o método de Tocher, mais uma vez, ficou com o resultado inferior. No que tange à diversidade fenotípica, os respectivos métodos geraram os seguintes coeficientes: 1,06, 1,06 e 1,06. Segundo este critério, houve empate entre os métodos utilizados. Assim, a coleção nuclear escolhida para representar os acessos de *C. baccatum* foi aquela estabelecida a partir da análise de correspondência simples, isto é, os acessos:

37, 52, 53, 285, 288, 1362, 1364, 362, 363, 781, 1360, 2633, 2653, 2935, 2948, 2958, 2980, 3236, 3347, 3656, 3699, 3750, 3797, 3803, 3817, 3833, 3851 e 3856.

4.3.3 Coleção nuclear para *Capsicum chinense*

Na coleção de *C. chinense* (307 acessos), considerando-se a análise de correspondência simples, a dispersão gráfica dos acessos nos três primeiros eixos fatoriais, que explicaram 34% da variabilidade nos dados, revelou a formação de dois grupos compactos, que concentraram praticamente todos os acessos (Figura 4.5). Assim, não se observaram acessos com divergência extrema, que tenham se afastado muito das duas nuvens principais.

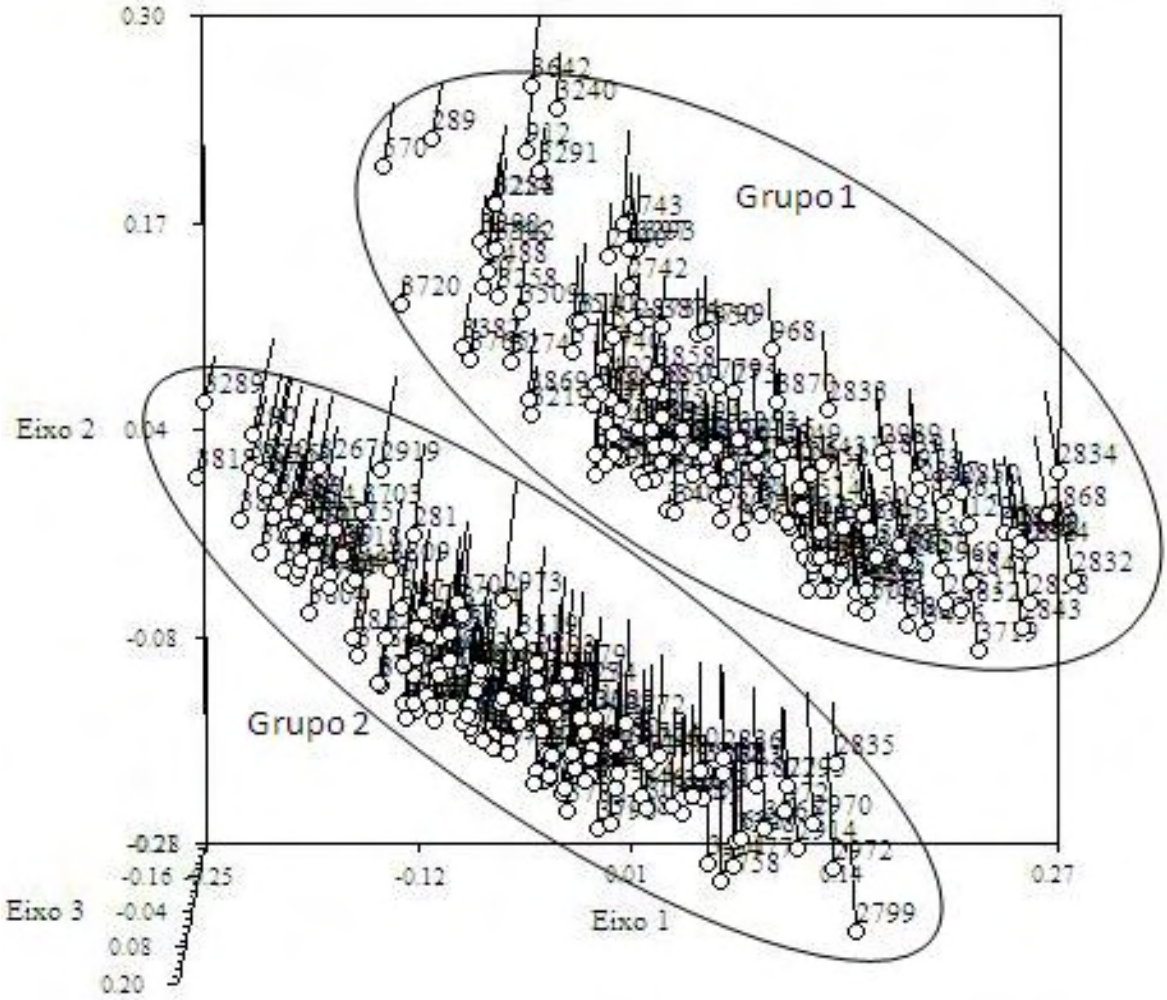

Figura 4.5. Dispersão de acessos da espécie *Capsicum chinense* nos três primeiros eixos fatoriais da análise de correspondência simples, aplicada à coleção de pimentas e pimentões da Embrapa Hortaliças.

Os acessos que mais se distanciaram na dispersão gráfica foram 2799 e 3642, que divergiram em 23 descritores. Suas diferenças mais marcantes estão em: aroma, que é baixo para o primeiro acesso e alto no segundo; espessura da parede do fruto, de 1-2 mm para 2799 e acima de 5 mm para 3642; formato do fruto, campanulada no acesso 2799 e triangular em 3642; peso do fruto, de 27-81 g em 2799 e até 1 g em 3642; e pungência, em que o acesso 2799 possui fruto doce e 3642 possui alta picância. Novamente, destacam-se caracteres do fruto como determinantes dos espectros fatoriais responsáveis pela maior discriminação dos acessos (Teixeira, 1996; Reifschneider, 2000; Sudré, 2006).

Os acessos que mais contribuíram para o primeiro eixo fatorial da análise de correspondência foram: 55, 59, 280, 290, 434, 599, 973, 1294, 2821, 2825, 2832, 2834, 2837, 2839, 2840, 2842, 2843, 2844, 2848, 2850, 2851, 2852, 2853, 2858, 2862, 2865, 2867, 2868, 2939, 2969, 3235, 3251, 3267, 3275, 3289, 3456, 3474, 3513, 3518, 3520, 3643, 3657, 3692, 3700, 3719, 3721, 3759, 3761, 3762, 3763, 3764, 3802, 3804, 3805, 3806, 3807, 3813, 3816, 3818, 3830 e 3844 (Tabela 4.4). Estes representaram cerca de 60% da variabilidade dos dados associada a este eixo multivariado. No segundo eixo, os acessos 60, 289, 436, 570, 573, 635, 679, 912, 968, 972, 2740, 2742, 2743, 2795, 2799, 2873, 2914, 2950, 2972, 3008, 3040, 3221, 3240, 3245, 3247, 3254, 3258, 3265, 3274, 3276, 3288, 3290, 3291, 3293, 3375, 3382, 3395, 3396, 3397, 3398, 3450, 3451, 3454, 3466, 3477, 3488, 3509, 3510, 3512, 3642, 3720, 3740, 3758, 3766, 3774, 3785, 3788, 3790, 3799, 3871 e 3874 foram os que apresentaram as maiores contribuições, representando, aproximadamente, 62% da variação dos dados neste eixo.

Tabela 4.4. Contribuição dos acessos da espécie *Capsicum chinense* para a variabilidade dos dados nos três primeiros eixos fatoriais da análise de correspondência, estimada a partir de 56 descritores qualitativos analisadas em 307 acessos da coleção de pimentas e pimentões da Embrapa Hortaliças.

Aces	Eixo	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3
55	0,89	0,02	0,01	2971	0,08	0,15	0,32	3454	0,04	0,93	1,69	3718	0,57	0,00	0,00
59	0,75	0,05	0,11	2972	0,41	0,83	0,59	3455	0,13	0,41	0,17	3719	1,25	0,07	0,05
60	0,16	0,61	3,01	2973	0,10	0,00	1,39	3456	0,87	0,07	0,07	3720	0,42	0,62	1,25
61	0,34	0,11	0,01	2975	0,02	0,07	0,70	3457	0,44	0,00	0,01	3721	0,78	0,01	0,28
275	0,14	0,29	1,09	3008	0,06	0,64	0,76	3458	0,26	0,16	0,21	3724	0,01	0,37	0,00
280	0,84	0,11	0,50	3040	0,01	0,62	0,19	3459	0,05	0,12	0,31	3725	0,00	0,49	0,21
281	0,37	0,04	0,28	3188	0,45	0,03	0,02	3460	0,05	0,49	0,73	3726	0,32	0,01	0,31
289	0,27	2,19	0,00	3219	0,06	0,23	0,23	3461	0,01	0,53	0,13	3740	0,00	0,63	0,00
290	0,93	0,29	0,54	3220	0,08	0,16	0,00	3464	0,41	0,00	0,00	3741	0,02	0,15	0,00
434	1,30	0,02	0,00	3221	0,04	0,62	0,01	3465	0,43	0,01	0,01	3748	0,28	0,28	0,00
436	0,00	0,69	0,57	3222	0,08	0,13	0,00	3466	0,19	0,70	2,62	3751	0,02	0,30	0,02
503	0,07	0,35	0,58	3233	0,18	0,07	0,55	3467	0,00	0,03	2,70	3758	0,10	1,04	0,15
570	0,47	2,13	0,01	3234	0,01	0,10	0,30	3468	0,08	0,27	0,01	3759	0,89	0,18	1,13
573	0,00	0,68	0,00	3235	0,92	0,18	0,55	3469	0,15	0,02	0,00	3760	0,34	0,35	0,46
578	0,32	0,13	2,19	3238	0,35	0,04	0,58	3471	0,03	0,04	0,03	3761	0,95	0,01	0,40
599	0,67	0,01	0,01	3239	0,29	0,15	0,10	3474	0,75	0,12	1,08	3762	0,84	0,15	1,24
600	0,27	0,03	0,11	3240	0,03	2,40	0,00	3475	0,53	0,00	0,09	3763	0,85	0,01	0,33
635	0,14	0,73	0,28	3245	0,04	0,66	0,16	3476	0,38	0,02	0,01	3764	0,76	0,09	1,36
670	0,32	0,10	0,03	3246	0,00	0,43	0,26	3477	0,12	0,95	0,18	3765	0,19	0,44	0,39
679	0,07	0,98	1,32	3247	0,01	0,90	0,14	3478	0,00	0,26	0,46	3766	0,12	1,26	0,09
779	0,11	0,51	0,03	3249	0,12	0,16	0,03	3480	0,18	0,03	0,08	3767	0,34	0,22	0,01
780	0,00	0,38	0,05	3250	0,01	0,31	0,34	3485	0,00	0,03	2,69	3768	0,46	0,18	0,03
912	0,07	2,55	0,31	3251	0,65	0,07	0,68	3486	0,54	0,01	0,02	3769	0,46	0,16	0,01
968	0,22	0,59	0,01	3252	0,06	0,11	0,03	3487	0,11	0,06	0,04	3770	0,34	0,07	0,06
969	0,09	0,09	0,23	3253	0,13	0,17	0,15	3488	0,14	1,31	0,44	3771	0,18	0,18	0,11
970	0,22	0,01	0,04	3254	0,12	1,51	0,05	3489	0,01	0,18	0,16	3772	0,11	0,38	0,00
971	0,21	0,02	0,30	3255	0,18	0,22	0,05	3490	0,00	0,23	0,00	3773	0,37	0,24	0,05
972	0,25	0,56	0,29	3256	0,00	0,13	0,16	3491	0,19	0,06	0,15	3774	0,02	0,56	0,14
973	0,79	0,12	0,01	3258	0,12	0,69	0,77	3492	0,00	0,37	0,09	3776	0,17	0,25	0,06
1294	1,10	0,03	0,01	3259	0,00	0,19	0,36	3493	0,01	0,22	0,03	3777	0,10	0,35	0,00
2740	0,00	1,19	0,06	3260	0,01	0,28	0,03	3494	0,04	0,18	0,21	3778	0,18	0,20	0,19
2742	0,00	1,22	0,36	3261	0,02	0,07	0,84	3495	0,35	0,00	0,00	3779	0,14	0,37	0,01
2743	0,00	1,75	0,84	3262	0,27	0,02	0,07	3501	0,57	0,03	0,05	3780	0,05	0,50	0,02
2795	0,25	0,57	0,00	3264	0,18	0,01	0,04	3502	0,03	0,00	1,76	3784	0,58	0,11	0,02
2799	0,49	1,49	0,02	3265	0,20	0,80	0,03	3506	0,01	0,14	0,19	3785	0,02	0,57	0,29
2803	0,05	0,03	1,06	3267	0,73	0,14	0,01	3507	0,35	0,00	0,00	3786	0,04	0,34	0,18
2821	0,77	0,05	0,07	3268	0,01	0,37	0,28	3508	0,55	0,01	0,03	3787	0,01	0,41	0,18
2822	0,07	0,38	0,48	3269	0,32	0,00	0,11	3509	0,08	0,74	0,10	3788	0,00	0,59	0,59
2823	0,11	0,39	0,45	3270	0,02	0,31	0,29	3510	0,01	0,82	0,00	3789	0,03	0,41	0,34
2824	0,25	0,24	0,70	3271	0,09	0,20	0,01	3511	0,16	0,21	0,27	3790	0,00	0,68	0,50
2825	0,64	0,00	0,01	3272	0,00	0,18	0,44	3512	0,11	1,46	0,65	3791	0,00	0,16	0,19

Tabela 4.4. (Continuação)

Aces	Eixo	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3
2832	1,67	0,00	0,09	3273	0,05	0,31	0,55	3513	0,78	0,00	0,05	3792	0,05	0,37	0,76
2833	0,38	0,37	0,10	3274	0,10	0,59	0,00	3514	0,33	0,07	0,05	3793	0,13	0,36	0,03
2834	1,60	0,15	0,10	3275	0,68	0,01	0,03	3517	0,14	0,13	0,01	3794	0,10	0,26	0,82
2835	0,41	0,51	0,12	3276	0,01	0,59	0,58	3518	0,64	0,01	0,70	3795	0,02	0,42	0,45
2836	0,11	0,40	0,04	3277	0,09	0,11	0,00	3519	0,08	0,10	0,19	3796	0,57	0,01	0,16
2837	0,76	0,15	0,59	3278	0,35	0,00	0,02	3520	1,05	0,13	0,00	3798	0,08	0,13	0,05
2838	0,27	0,18	1,47	3279	0,02	0,11	0,06	3521	0,03	0,04	3,21	3799	0,07	0,65	0,05
2839	0,63	0,17	0,04	3283	0,09	0,03	0,25	3523	0,21	0,49	1,96	3800	0,10	0,02	0,06
2840	1,45	0,01	0,00	3284	0,02	0,12	0,05	3524	0,00	0,48	0,21	3801	0,60	0,01	0,22
2842	1,13	0,00	0,06	3287	0,62	0,01	0,05	3527	0,17	0,24	0,18	3802	0,77	0,02	0,15
2843	1,47	0,01	0,27	3288	0,10	1,67	0,29	3531	0,03	0,15	0,57	3804	0,79	0,06	0,43
2844	1,54	0,01	0,00	3289	1,41	0,26	0,26	3639	0,01	0,16	0,34	3805	0,74	0,01	0,07
2848	1,40	0,02	0,04	3290	0,15	1,13	0,15	3640	0,22	0,17	0,10	3806	0,93	0,03	0,00
2849	0,25	0,23	0,15	3291	0,05	1,75	0,18	3641	0,04	0,22	0,22	3807	0,81	0,02	0,05
2850	1,03	0,07	0,02	3293	0,00	1,39	0,14	3642	0,05	2,75	0,21	3808	0,40	0,00	0,17
2851	1,00	0,00	0,13	3338	0,00	0,35	0,04	3643	0,69	0,02	0,05	3809	0,39	0,00	0,00
2852	1,10	0,15	2,78	3345	0,27	0,00	0,48	3644	0,05	0,16	0,04	3812	0,56	0,05	0,00
2853	1,51	0,00	0,00	3346	0,55	0,03	0,09	3651	0,02	0,19	0,00	3813	1,08	0,00	0,14
2854	0,39	0,02	0,01	3373	0,26	0,20	0,04	3652	0,22	0,05	0,47	3816	1,08	0,01	0,16
2858	1,49	0,00	1,21	3375	0,03	0,57	0,21	3653	0,10	0,20	0,00	3818	1,51	0,04	0,65
2862	1,50	0,02	0,05	3378	0,00	0,27	0,00	3657	0,89	0,02	1,40	3826	0,00	0,26	0,81
2865	0,98	0,05	0,01	3382	0,20	0,57	0,08	3659	0,49	0,05	1,52	3830	0,78	0,01	0,00
2867	0,75	0,00	0,08	3383	0,36	0,09	0,05	3690	0,34	0,12	0,01	3844	0,72	0,01	0,46
2868	1,57	0,05	0,06	3384	0,25	0,10	0,10	3692	0,85	0,00	0,02	3845	0,00	0,18	0,00
2872	0,16	0,43	0,21	3386	0,27	0,16	0,22	3693	0,54	0,00	0,11	3846	0,37	0,29	0,02
2873	0,00	0,66	0,09	3395	0,01	0,76	0,44	3694	0,16	0,20	0,16	3850	0,01	0,38	0,07
2909	0,15	0,09	0,09	3396	0,01	0,85	1,01	3695	0,13	0,21	0,60	3857	0,00	0,36	0,05
2914	0,31	1,25	0,42	3397	0,00	1,42	0,19	3700	0,82	0,16	0,56	3858	0,02	0,50	0,02
2919	0,36	0,16	0,33	3398	0,13	0,80	0,19	3701	0,00	0,12	0,06	3859	0,02	0,33	0,10
2939	0,63	0,14	0,12	3429	0,07	0,38	0,46	3702	0,45	0,05	0,01	3864	0,02	0,35	0,21
2945	0,12	0,30	0,25	3431	0,39	0,06	0,87	3703	0,50	0,03	0,00	3869	0,06	0,39	0,02
2950	0,06	0,59	0,24	3447	0,37	0,01	0,05	3705	0,04	0,54	0,00	3870	0,22	0,31	0,04
2952	0,36	0,10	0,04	3449	0,00	0,01	2,56	3706	0,03	0,27	0,58	3871	0,05	0,56	0,08
2967	0,00	0,42	0,09	3450	0,03	0,98	1,38	3713	0,01	0,30	0,01	3874	0,02	0,67	0,12
2969	0,93	0,00	0,03	3451	0,08	1,09	0,00	3714	0,01	0,01	2,27	3875	0,19	0,22	0,19
2970	0,33	0,49	1,36	3452	0,35	0,01	0,01	3717	0,28	0,07	0,03	–	–	–	–

Da análise desses resultados decorre a recomendação de que alguns destes acessos estejam representados na coleção nuclear associada a essa espécie. Assim, foi realizada a amostragem aleatória proporcional dentro dos dois grupos formados, o que resultou na seguinte relação de acessos (20% do tamanho da coleção inicial), para compor

a coleção nuclear de *C. chinense*, segundo este método: 59, 60, 275, 290, 570, 578, 780, 2742, 2799, 2822, 2832, 2833, 2852, 2867, 2909, 2919, 2952, 2969, 2972, 2973, 3188, 3219, 3238, 3239, 3253, 3254, 3259, 3260, 3261, 3265, 3267, 3275, 3289, 3375, 3395, 3396, 3449, 3450, 3456, 3457, 3459, 3460, 3466, 3487, 3489, 3492, 3502, 3519, 3521, 3642, 3657, 3659, 3703, 3720, 3760, 3767, 3791, 3792, 3813, 3818, 3870 e 3874.

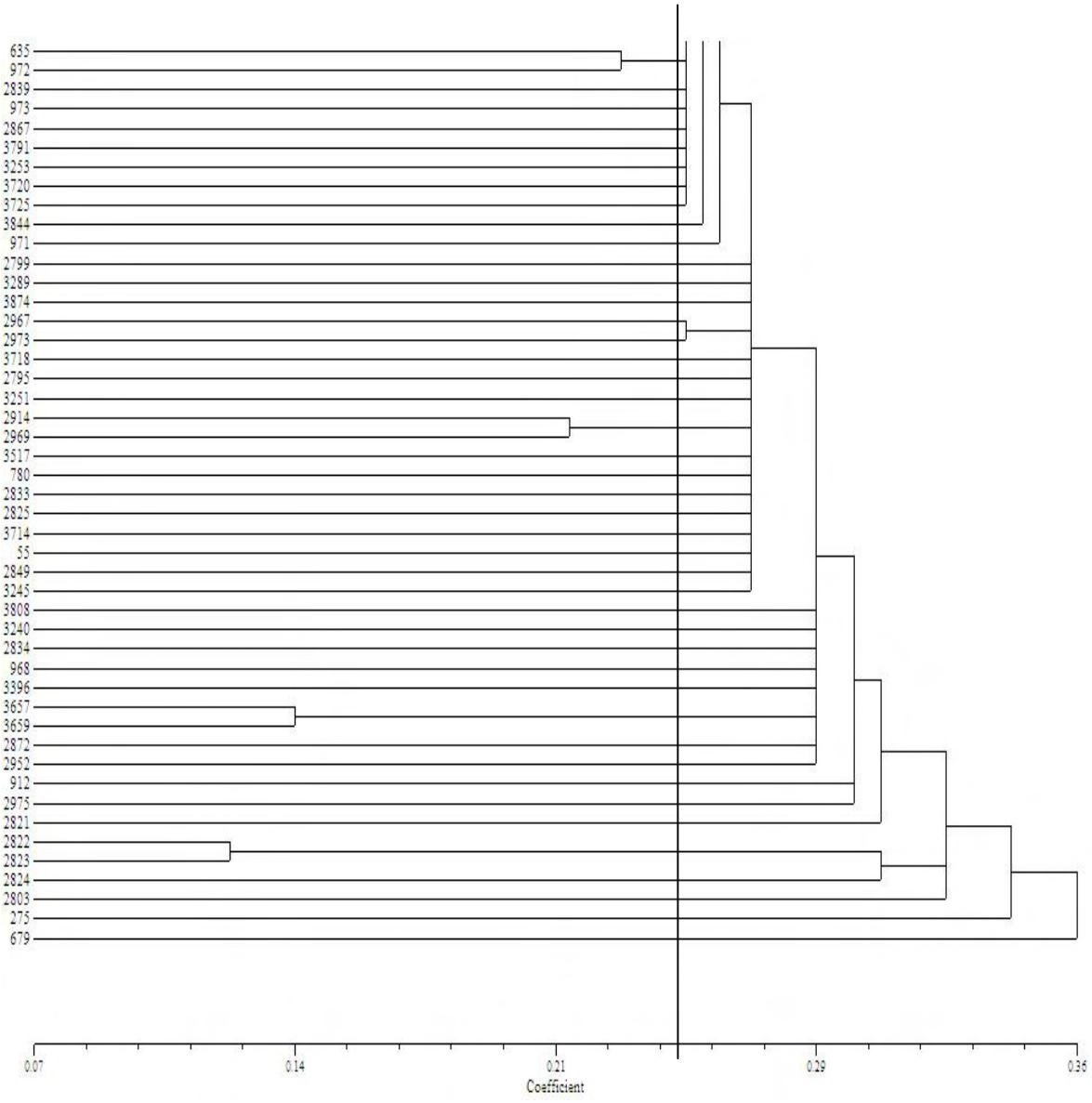

Figura 4.6. Dendrograma ilustrativo do padrão de dissimilaridade entre 307 acessos de *Capsicum chinense* da coleção de pimentas e pimentões da Embrapa Hortaliças estimado pelo método hierárquico do vizinho mais próximo, com base nas distâncias para variáveis multicategóricas obtidas a partir de 56 descritores qualitativos (continua).

Figura 4.6. (Continuação)

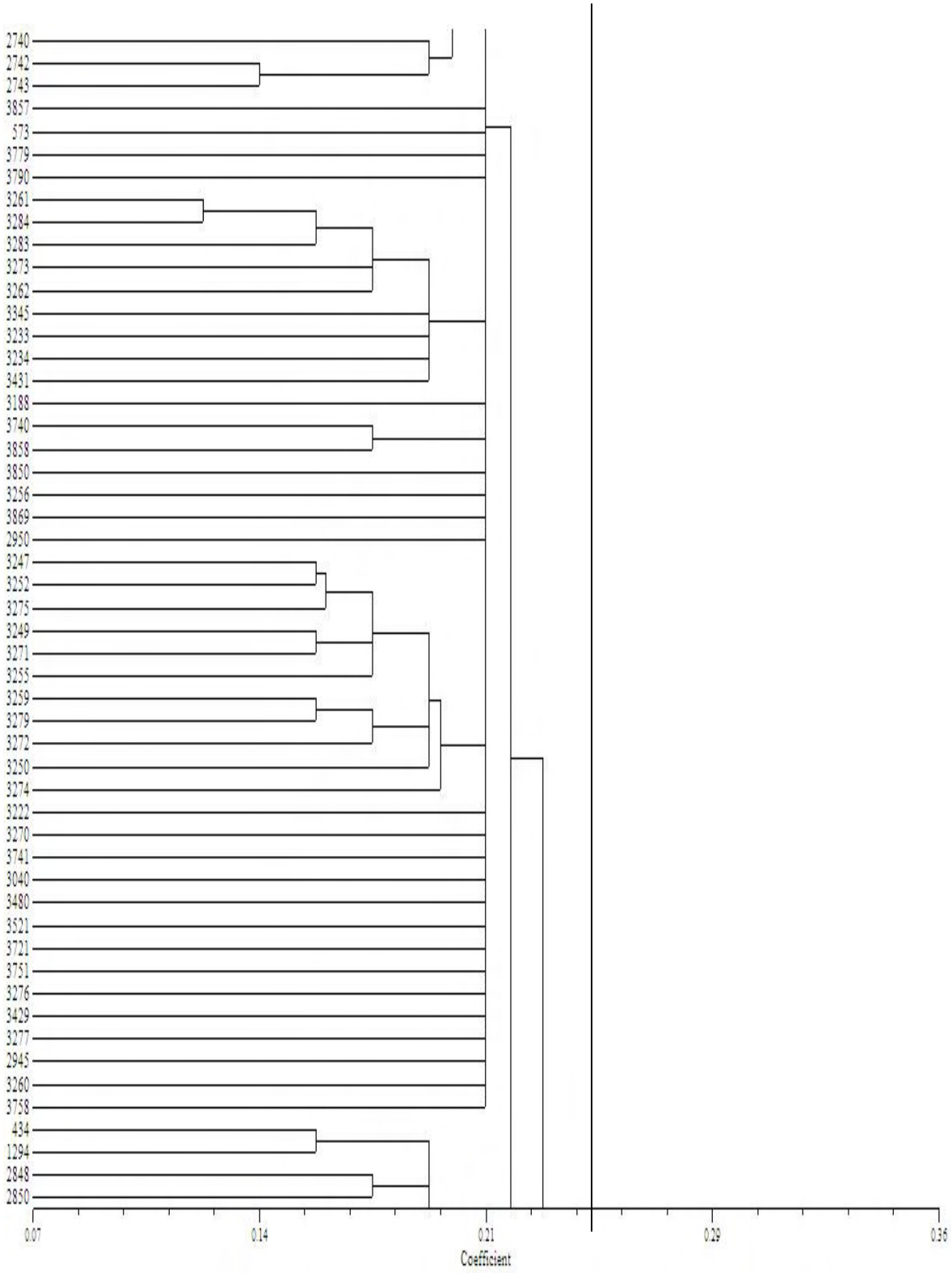

Figura 4.6. (Continuação)

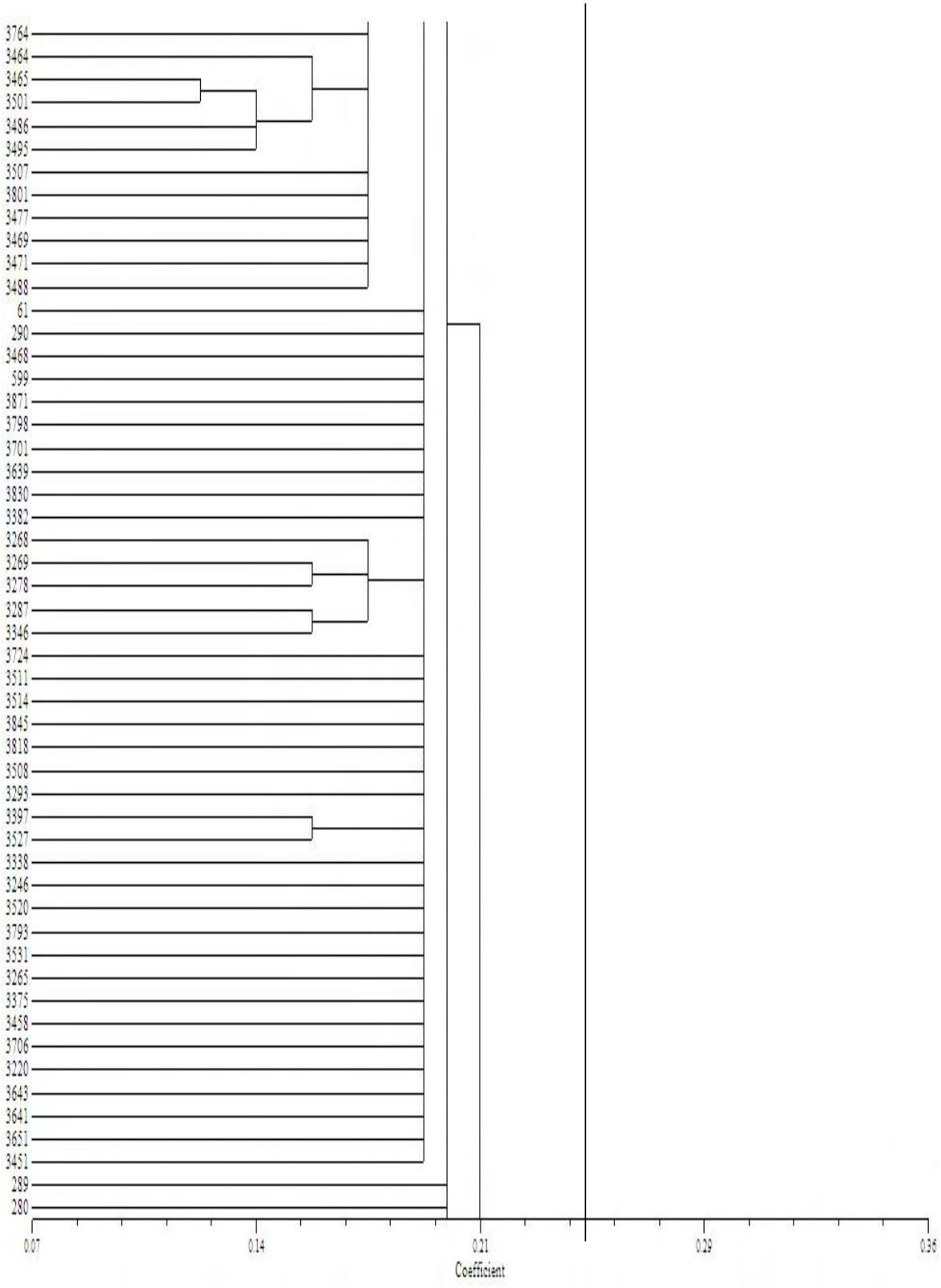

Figura 4.6. (Continuação)

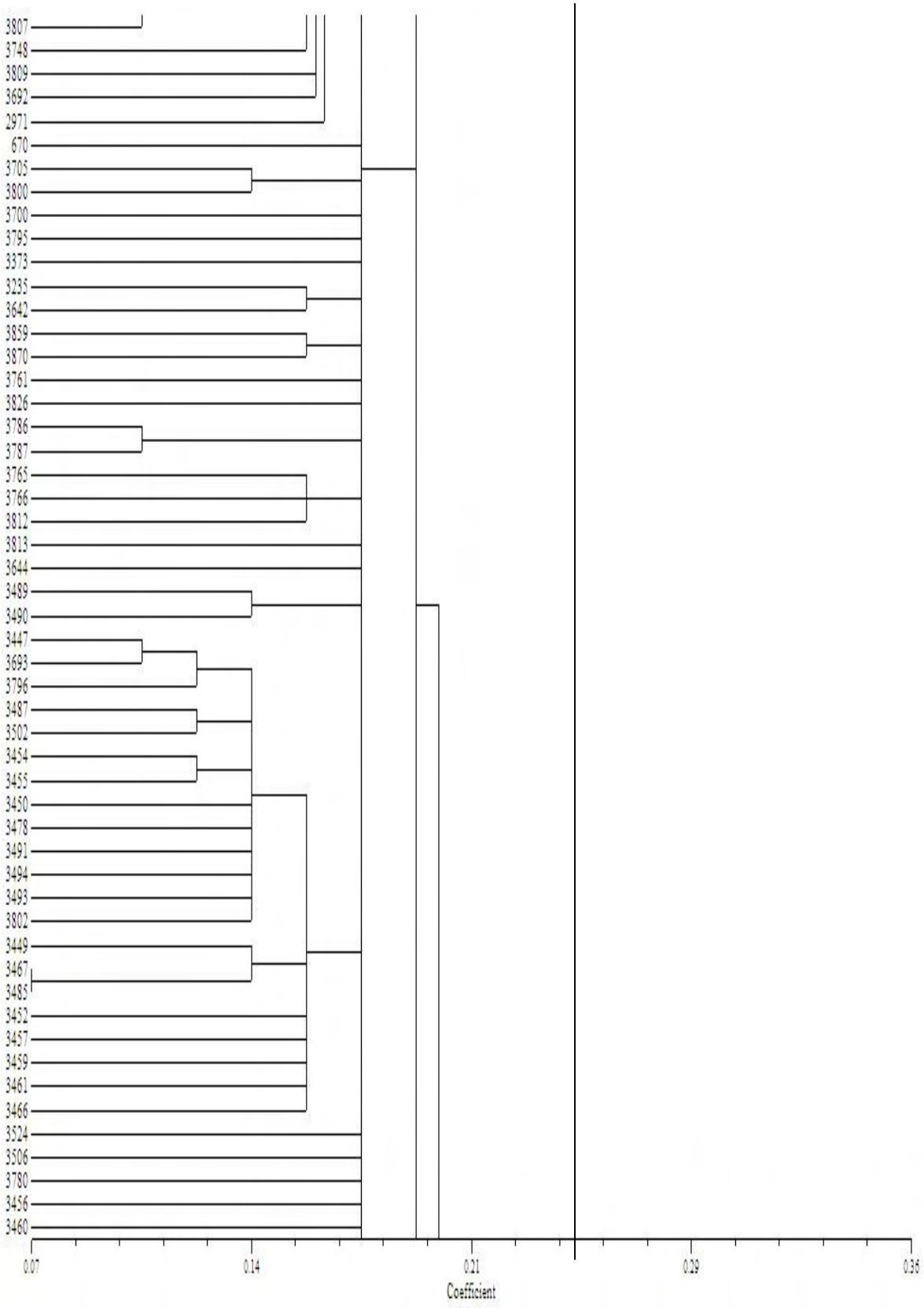

Figura 4.6. (Continuação)

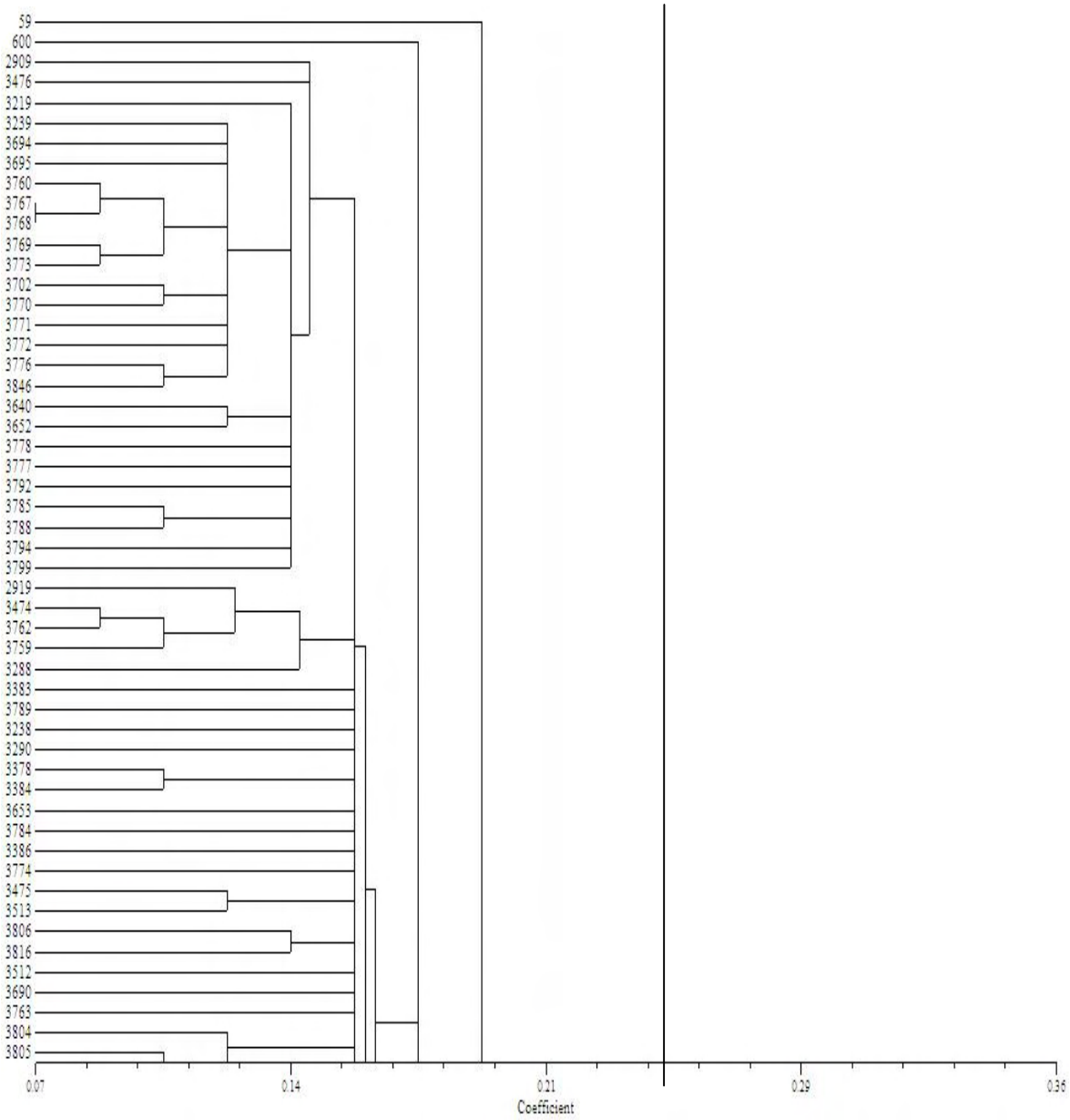

Figura 4.6. (Continuação)

Pela análise de agrupamento usando o método do vizinho mais próximo e adotando-se a intensidade de seleção de 20%, obtiveram-se 54 grupos de similaridade, a partir de um corte na distância de ligação 0,23 do dendrograma (Figura 4.6): um grupo maior com 242 acessos, um grupo com quatro acessos, um grupo com três acessos, cinco grupos com dois acessos, e 46 grupos de acessos isolados. Assim, dos 62 acessos a comporem a coleção nuclear, restaram dezesseis para serem sorteados dentro dos grupos com dois ou mais acessos. Destes, sete acessos foram aleatoriamente escolhidos nos grupos com até quatro acessos – um em cada, e nove foram sorteados no grupo maior. Enfim, a coleção nuclear proposta para *C. chinense,* segundo este método, foi composta pelos

seguintes acessos: 55, 275, 281, 289, 600, 679, 780, 912, 968, 971, 972, 973, 2795, 2799, 2803, 2821, 2822, 2823, 2824, 2825, 2833, 2834, 2835, 2838, 2839, 2849, 2865, 2867, 2872, 2914, 2952, 2967, 2969, 2973, 2975, 3008, 3240, 3245, 3251, 3253, 3289, 3291, 3395, 3396, 3466, 3485, 3510, 3517, 3518, 3657, 3659, 3714, 3718, 3719, 3720, 3725, 3763, 3791, 3808, 3830, 3844 e 3874. Destes, 31% foram também selecionados para compor a coleção estabelecida pela análise de correspondência, o que corresponde a 16 acessos em comum nas duas coleções nucleares.

Por este método, pôde-se verificar, ainda, que dois pares de acessos (3767 e 3768) e (3467 e 3485) mostraram similaridades tão elevadas, que revelam suspeita de também serem duplicatas. Assim, justificam-se análises genéticas do tipo *fingerprint*, para a confirmação deste fato via marcadores moleculares, o que também reforça a importância da caracterização molecular das coleções de germoplasma.

Pelo método de Tocher, com critério de aglomeração inverso, considerando-se a intensidade de amostragem estabelecida (20%), a coleção nuclear de *C. chinense* seria formada pelos seguintes acessos: 434, 436, 503, 570, 573, 599, 912, 972, 2742, 2743, 2795, 2821, 2822, 2835, 2872, 2919, 2952, 3239, 3240, 3249, 3250, 3251, 3255, 3260, 3262, 3271, 3272, 3275, 3276, 3277, 3289, 3345, 3386, 3450, 3451, 3460, 3469, 3476, 3477, 3485, 3486, 3490, 3508, 3524, 3531, 3694, 3702, 3713, 3717, 3726, 3762, 3764, 3767, 3779, 3790, 3794, 3804, 3808, 3818, 3826, 3830 e 3845. Esta coleção teve concordância de 21% com a coleção nuclear obtida por análise de correspondência e de 23% com a coleção nuclear obtida pelo método de agrupamento do vizinho mais próximo.

Obtidas as coleções nucleares a partir dos três métodos, verificou-se, então, os coeficientes de similaridade (CR%) e de diversidade fenotípica (DF), associados a estas coleções. Para a análise de correspondência, o índice CR% entre a coleção nuclear obtida e sua correspondente coleção inicial foi de 90,8%; para o método de agrupamento pelo vizinho mais próximo, 92,5%; e para o método de Tocher, 88,6%. Logo, mais uma vez, os três métodos revelaram níveis de similaridade satisfatórios para a constituição de coleções nucleares. Contudo, a vantagem, embora pequena, voltou a favorecer o método de agrupamento pelo vizinho mais próximo, e a desvantagem relativa novamente recaiu sobre o método de Tocher.

Os coeficientes de diversidade fenotípica (DF), considerando-se as coleções nucleares obtidas pela análise de correspondência, pelo método de agrupamento do vizinho mais próximo e pelo método de Tocher foram, respectivamente, 1,07, 1,19 e 1,04.

Novamente os índices revelam propostas de coleções satisfatórias, pois a distribuição dos acessos selecionados nas categorias de variáveis observadas é mais equilibrada nestas coleções do que na coleção base (valores de DF>1,0). Conforme já comentado, isso vem atender uma expectativa presente no estabelecimento de coleções nucleares, que é o aumento das variâncias nestas coleções relativamente às suas coleções base (van Hintum et al., 2000). A ordem desses coeficientes, entretanto, mais uma vez favorece o método do vizinho mais próximo, sobretudo em relação ao método de Tocher. Esse resultado tende a contrariar a recomendação de Vasconcelos et al. (2007), em favor da aplicação deste método; embora o propósito do seu estudo não tenha sido o de comparar métodos, mas sim, avaliar a influência da intensidade de amostragem na qualidade das coleções nucleares obtidas a partir do emprego desse método.

Dessa forma, novamente conclui-se que a coleção nuclear gerada a partir do método de agrupamento do vizinho mais próximo deve ser a escolhida para representar os acessos da coleção inicial de *C. chinense*, isto é, os acessos: 55, 275, 281, 289, 600, 679, 780, 912, 968, 971, 972, 973, 2795, 2799, 2803, 2821, 2822, 2823, 2824, 2825, 2833, 2834, 2835, 2838, 2839, 2849, 2865, 2867, 2872, 2914, 2952, 2967, 2969, 2973, 2975, 3008, 3240, 3245, 3251, 3253, 3289, 3291, 3395, 3396, 3466, 3485, 3510, 3517, 3518, 3657, 3659, 3714, 3718, 3719, 3720, 3725, 3763, 3791, 3808, 3830, 3844 e 3874.

4.3.4 Coleção nuclear para *Capsicum frutescens*

No caso da coleção de *C. frutescens*, com apenas 60 acessos, a análise de correspondência simples, nos seus três primeiros eixos fatoriais, capturou 44% da variabilidade dos dados. O gráfico de dispersão fatorial dos indivíduos, associado à análise, revelou a formação de dois grupos principais de acessos (Figura 4.7), identificados no gráfico à direita e à esquerda do segundo eixo fatorial. Dentro destes grupos, visualizam-se acessos mais afastados do aglomerado central, que, de algum modo, deveriam ser priorizados na composição da coleção nuclear, hajam vista suas divergências relativas. No grupo à esquerda, os acessos que mais se dispersaram são 3822, 3645 e 2864 (na parte superior do gráfico) e 3835, 2687 e 3697, na parte inferior; e no grupo à direita, 3649, 3349, 2871, 2846, 2870, mais acima na figura, e 3000, 3715, 3716 e 2941, mais em baixo. Esses acessos estiveram, portanto, entre aqueles que mais contribuíram para os primeiros eixos fatoriais da análise.

Especificamente, no primeiro eixo destacam-se 3000, 2846, 2870, 2871, 3448, 3649, 3349, 3746, 596, 2744, 3257, 3696, que representaram mais de 37% da variabilidade associada a esse eixo (Tabela 4.5). No segundo, os acessos que mais contribuíram foram 3822, 3649, 3349, 3835, 3645, 2941, 3716, 2864, 3000, 2687, 3286, 3697, com mais de 68% da variabilidade relacionada a este fator.

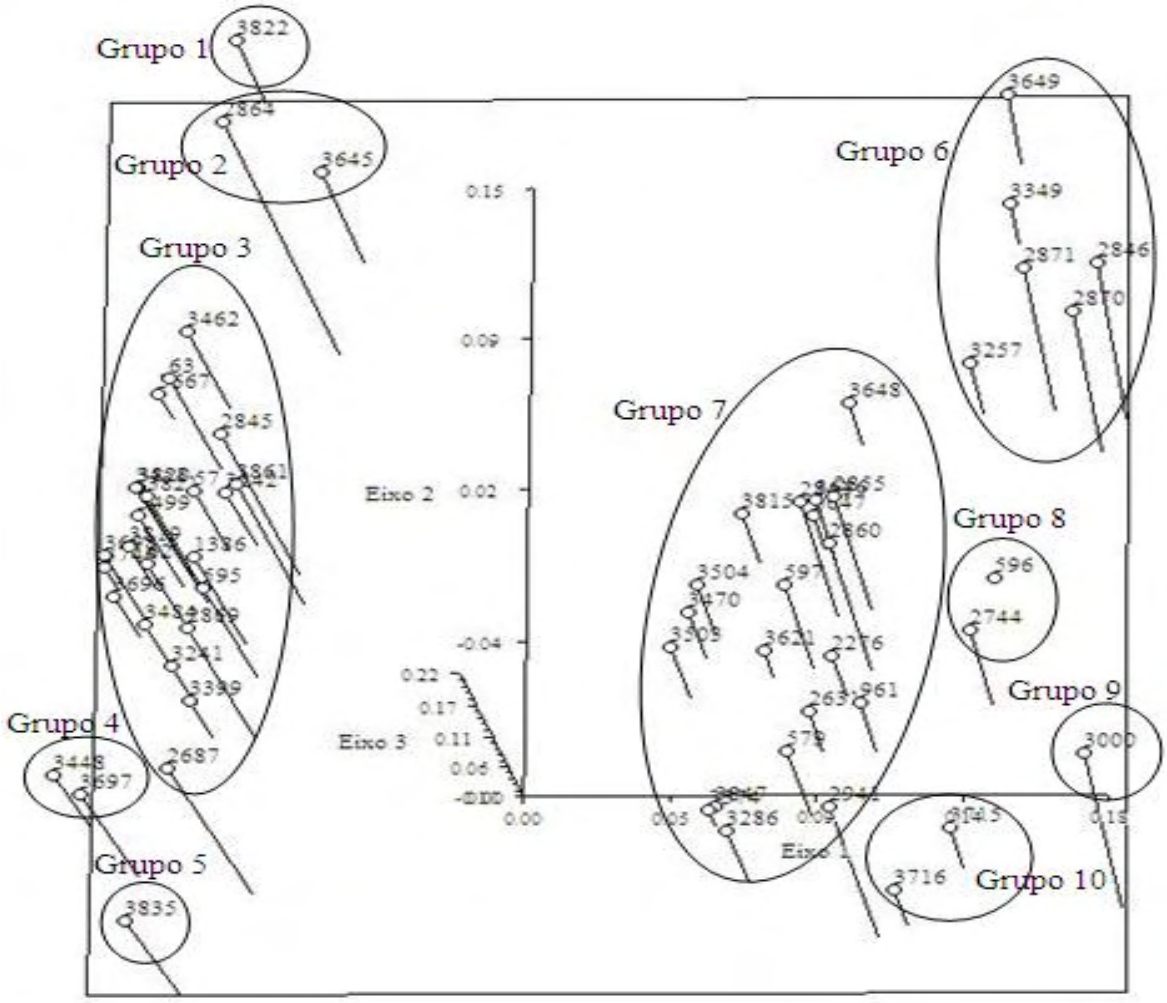

Figura 4.7. Dispersão de acessos da espécie *Capsicum frutescens* nos três primeiros eixos da análise de correspondência simples, aplicada à coleção de pimentas e pimentões da Embrapa Hortaliças.

Os acessos que se revelaram mais divergentes entre si por essa análise foram 3822 e 3835, que se diferiram em quatorze descritores, com destaque para suas diferentes procedências, em que 3822 é do sul do Brasil e 3835 oriundo de coletas nos EUA e Canadá. Outro par divergente foi 3000 e 3649, que mostraram diferenças em dezenove descritores, sendo também a procedência o descritor mais marcante (3000 é originário da Ásia e 3649 do nordeste do Brasil). Outros dois acessos discrepantes foram 3000 e 3448,

que se divergiram em dezesseis descritores, tendo a procedência também como descritor principal (3000 é da Ásia e 3448 do norte do Brasil). Essa constatação sugere que os acessos de *C. frutescens,* da coleção da Embrapa Hortaliças, podem ser discriminados morfologicamente a partir da informação de seus locais de origem; o que revela também provável efeito evolutivo decorrente de isolamento geográfico. Enfim, para garantir uma boa representação dos acessos, foram identificados dez grupos de similaridade, dentro dos quais foram amostrados acessos, aleatoriamente, considerando-se a intensidade estabelecida (30%) e garantindo-se, pelo menos, um acesso por grupo. Logo, os acessos selecionados por esse método, para a composição da coleção nuclear de *C. frutescens*, foram os seguintes: 63, 596, 961, 1386, 2276, 2846, 2864, 2941, 3000, 3399, 3647, 3648, 3649, 3696, 3697, 3716, 3822 e 3835.

Tabela 4.5. Contribuição percentual dos acessos da espécie *Capsicum frutescens* para a variabilidade dos dados nos três primeiros eixos fatoriais da análise de correspondência, estimada a partir de 56 descritores qualitativos analisadas em sessenta acessos da coleção de pimentas e pimentões da Embrapa Hortaliças.

Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3	Aces	Eixo1	Eixo2	Eixo3
57	1,52	0,28	0,26	2846	4,20	1,88	4,17	3399	1,70	0,61	1,12	3667	2,18	2,08	2,40
62	1,93	0,00	0,10	2855	1,34	0,01	1,45	3448	3,05	1,97	0,22	3696	2,41	0,01	0,93
63	1,64	1,17	0,16	2860	1,31	0,11	2,71	3453	1,77	0,02	1,64	3697	2,35	2,90	0,34
579	0,89	1,78	0,01	2864	0,77	3,70	16,88	3462	1,51	2,18	0,01	3698	2,38	0,00	0,03
595	1,28	0,13	0,31	2869	1,25	0,59	1,55	3470	0,30	0,06	0,70	3715	2,23	2,82	0,69
596	2,71	0,09	5,21	2870	3,95	1,35	2,98	3484	1,80	0,25	0,02	3716	1,73	4,42	1,07
597	0,92	0,10	0,13	2871	3,48	2,22	3,11	3499	1,98	0,07	0,00	3746	2,75	0,06	2,61
961	1,46	0,83	0,49	2941	1,37	4,50	5,08	3503	0,24	0,28	0,41	3815	0,54	0,17	0,57
1386	1,38	0,02	0,24	3000	4,21	3,70	8,74	3504	0,34	0,01	0,47	3819	2,30	0,07	0,94
2276	1,21	0,28	0,93	3241	1,78	0,30	0,95	3621	0,66	0,15	2,13	3820	1,92	0,11	0,19
2631	1,03	0,84	1,04	3242	1,24	0,29	0,42	3645	0,56	6,18	0,04	3821	1,75	0,05	0,49
2687	1,28	3,42	4,30	3257	2,43	2,06	0,64	3646	1,05	0,18	0,15	3822	1,31	13,14	0,45
2744	2,62	0,34	0,03	3286	0,53	3,30	0,23	3647	0,98	0,27	1,68	3835	1,88	6,30	0,22
2841	1,08	0,00	1,55	3349	2,82	6,82	1,38	3648	1,25	1,51	1,05	3847	0,35	1,66	5,39
2845	0,93	0,12	3,69	3374	0,36	2,03	3,15	3649	2,91	10,21	0,18	3861	0,92	0,02	1,95

Pelo método de agrupamento do vizinho mais próximo, após um corte no dendrograma obtido, na distância de ligação de 0,17 (Figura 4.8), observou-se a formação de quinze grupos de similaridade, sendo o agrupamento maior com quarenta acessos, seguido de um grupo com seis, outro com dois acessos e doze grupos com acessos isolados. Os grupos menores correspondem, em grande parte, aos acessos mais afastados das nuvens de pontos observadas na Figura 4.7, revelando certa concordância nos

resultados deste método com os da análise de correspondência. A partir do dendrograma (Figura 4.8) observou-se, também, uma possível duplicidade entre os acessos 3241 e 3242, que se agruparam em distância mínima.

Para composição da coleção nuclear, segundo este método, amostrou-se aleatoriamente quatro acessos no grupo maior e um acesso dentro de cada grupo menor, selecionando, assim, dezoito acessos (30% de sessenta acessos). Assim, os acessos amostrados, segundo esta abordagem, para representar a coleção de *C. frutescens* foram: 63, 1386, 2276, 2687, 2744, 2846, 2864, 2871, 2941, 3000, 3453, 3499, 3645, 3649, 3716, 3815, 3821 e 3861. Destes, 50% foram também selecionados para compor a coleção estabelecida pela análise de correspondência, o que corresponde a nove acessos em comum entre as duas coleções nucleares.

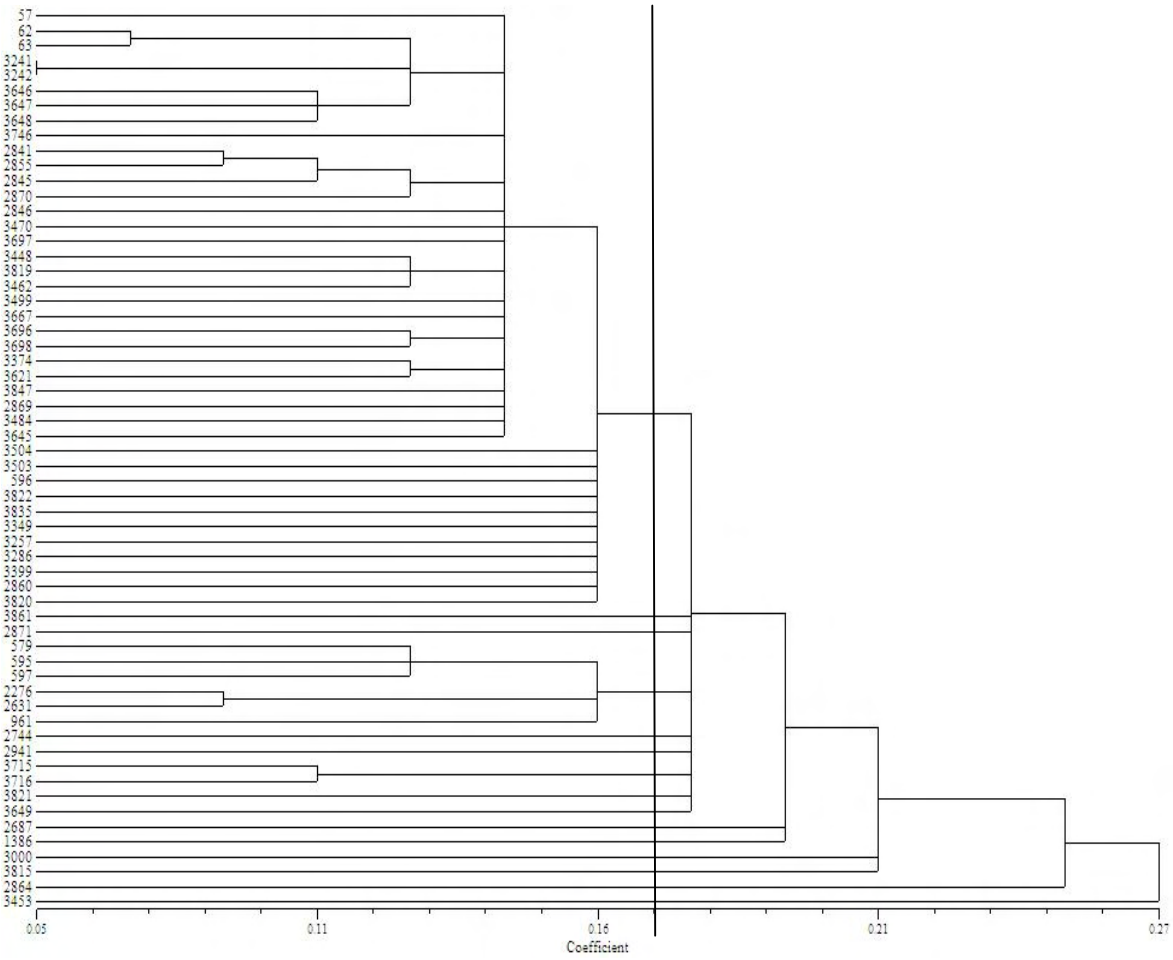

Figura 4.8. Dendrograma ilustrativo do padrão de dissimilaridade entre sessenta acessos de *Capsicum frutescens* da coleção de *Capsicum* spp. da Embrapa Hortaliças, estimado pelo método hierárquico do vizinho mais próximo com base nas distâncias para variáveis multicategóricas, obtidas a partir de 56 descritores qualitativos.

A partir do método de otimização de Tocher, com critério de aglomeração inverso, e considerando-se a intensidade de amostragem de 30%, os seguintes acessos foram selecionados para compor a coleção nuclear de *C. frutescens*: 595, 2846, 2864, 2941, 3257, 3462, 3645, 3647, 3648, 3667, 3697, 3716, 3746, 3815, 3819, 3820, 3821 e 3861. Esta coleção teve concordância de 39% com a coleção obtida por análise de correspondência e 44% com a coleção nuclear obtida pelo método de agrupamento do vizinho mais próximo.

Os coeficientes de similaridade (CR%) entre as coleções nucleares obtidas a partir dos três métodos empregados e a coleção completa de *C. frutescens* foram 90,2%, 94,3% e 84,6%, para a análise de correspondência, o agrupamento pelo vizinho mais próximo e o método de Tocher invertido, respectivamente. Os respectivos coeficientes de diversidade fenotípica (DF) foram 1,06, 1,16 e 1,04. Esses resultados revelam, mais uma vez, a eficácia dos três métodos utilizados para a composição da coleção nuclear, pois todos mostraram CR% acima de 80% e valores de DF superiores à unidade. Isso mostra que as coleções propostas preservam, no mínimo, 85% das categorias de variáveis observadas na coleção base, com uma distribuição de acessos nessas categorias tão equilibrada ou mais do que a coleção base. Infere-se, portanto, que os alelos responsáveis pela manifestação dessas categorias (fenótipos) são, por conseguinte, também preservados na coleção nuclear. A recomendação de Brown (1989) é que a representação desses alelos na coleção nuclear seja maximizada, respeitando-se o número de acessos considerado aceitável para essa coleção. Segundo Allard (1992), os alelos não-neutros que são selecionados natural ou artificialmente para adaptação a específicas condições ecológicas ou agrícolas (grupo dos alelos comuns e localizados) são potencialmente mais relevantes para os programas de melhoramento.

Por outro lado, evidencia-se também a vantagem relativa o método de agrupamento pelo vizinho mais próximo, sempre com coeficientes superiores nos dois critérios de validação utilizados. Também reitera-se, embora em menor nível, certa desvantagem relativa do método de Tocher, inclusive em relação à análise de correspondência simples. Segundo o critério DF, estes métodos praticamente empataram, mas em relação ao critério CR%, a coleção obtida pelo método de Tocher preservou, em média, 6% menos categorias do que a coleção obtida via análise de correspondência. Em razão disso, novamente, optou-se pela coleção nuclear obtida pelo método do vizinho mais próximo, para representar os acessos de *C. frutescens*, a saber: 63, 1386, 2276, 2687, 2744,

2846, 2864, 2871, 2941, 3000, 3453, 3499, 3645, 3649, 3716, 3815, 3821 e 3861. Entretanto, como já mencionado, essa menor eficiência do método de Tocher invertido poderá ser investigada, por exemplo, por meio de simulação, o que se justifica em razão das propriedades do método, no qual os acessos não são aleatoriamente escolhidos, como nos demais utilizados no presente trabalho, mas sim por um critério de divergência máxima (Vasconcelos, 2007).

4.3.5 Coleção nuclear para *Capsicum* spp.

Para a formação da coleção nuclear do banco de germoplasma de pimentas e pimentões da Embrapa Hortaliças, a estratégia adotada foi reunir as coleções nucleares propostas para os estratos principais inicialmente estabelecidos, isto é, as quatro espécies cultivadas de *Capsicum*. Esse agrupamento permite garantir que a coleção nuclear proposta tenha uma quantidade mínima e representativa de acessos de cada sub-coleção de espécie (*C. annuum*, *C. baccatum*, *C. chinense* e *C. frutescens*), preservando-se a variabilidade dentro da coleção genérica. É importante salientar que essa coleção poderá ser incrementada com novos acessos, de acordo com as necessidades dos melhoristas, e, também, para representar o aumento da variabilidade para novas características de interesse. Nesse sentido, é importante destacar o fato de que as espécies silvestres de *Capsicum* não foram consideradas neste estudo, embora esses materiais sejam também importantes fontes de diversidade genética a ser levadas em conta, futuramente, para o enriquecimento da coleção. De qualquer modo, não se deve esquecer a recomendação de van Hintum et al. (2000), de que é preferível ter uma coleção pequena, mas bem utilizada, do que uma grande, porém raramente requisitada.

Diante disso, a coleção nuclear final, proposta para o banco de germoplasma de *Capsicum* spp. da Embrapa Hortaliças, é formada pelos seguintes acessos: 184, 188, 674, 678, 685, 695, 699, 706, 715, 724, 733, 734, 749, 990, 991, 1376, 2175, 2176, 2277, 2278, 2683, 2684, 2688, 2725, 2769, 2785, 2802, 2817, 2905, 2923, 2926, 2942, 2953, 2955, 2988, 3400, 3498, 3595, 3707, 3722, 3745 e 3836, de *C. annuum*; 37, 52, 53, 285, 288, 362, 363, 781, 1360, 1362, 1364, 2633, 2653, 2935, 2948, 2958, 2980, 3236, 3347, 3656, 3699, 3750, 3797, 3803, 3817, 3833, 3851 e 3856, de *C. baccatum*; 55, 275, 281, 289, 600, 679, 780, 912, 968, 971, 972, 973, 2795, 2799, 2803, 2821, 2822, 2823, 2824, 2825, 2833, 2834, 2835, 2838, 2839, 2849, 2865, 2867, 2872, 2914, 2952, 2967, 2969, 2973, 2975, 3008, 3240, 3245, 3251, 3253, 3289, 3291, 3395, 3396, 3466, 3485, 3510, 3517, 3518,

3657, 3659, 3714, 3718, 3719, 3720, 3725, 3763, 3791, 3808, 3830, 3844 e 3874, de *C. chinense*; e 63, 1386, 2276, 2687, 2744, 2846, 2864, 2871, 2941, 3000, 3453, 3499, 3645, 3649, 3716, 3815, 3821 e 3861, de *C. frutescens*. Portanto, a coleção nuclear final fica composta por 150 acessos, o que corresponde a 17% da coleção inicial, situando-se na faixa de intensidade de amostragem comumente empregada nesse tipo de estudo (Brown, 1989; Brown & Spillane, 1999; Oliveira, 2007).

Os coeficientes de similaridade (CR%) e de diversidade fenotípica (DF) na coleção nuclear final (*Capsicum* spp.) são de 95,3% e 1,13, respectivamente. Verifica-se, assim, que a coleção proposta preserva mais de 95% das categorias de variáveis observadas na coleção completa, com uma distribuição de acessos nessas categorias mais equilibrada do que na coleção base.

4.4 CONCLUSÕES

- Todos os métodos avaliados mostraram-se eficazes no estabelecimento de coleções nucleares para *Capsicum* spp., com vantagem relativa para o método de agrupamento hierárquico pelo vizinho mais próximo, associado a análise de variáveis multicategóricas. O método de Tocher com critério de aglomeração inverso produziu coleções nucleares de menor similaridade e diversidade fenotípica com as respectivas coleções base.
- Análise de correspondência simples, cuja aplicação nesse tipo de estudo ainda não é descrita, apresenta-se como alternativa útil para o estabelecimento de coleções nucleares a partir de bancos de germoplasma cujos acessos encontram-se caracterizados segundo descritores qualitativos.
- As coleções nucleares obtidas pela análise de correspondência e pelo método de agrupamento hierárquico do vizinho mais próximo foram as que apresentaram maior nível de concordância (maior número de acessos comuns).
- Os pares de acessos 333 e 2912, 2957 e 2958, 3851 e 3853, 3767 e 3768, 3467 e 3485, 3241 e 3242, devido à alta similaridade, podem tratar-se de duplicatas. É importante a confirmação dessas duplicações a partir de análises específicas, via marcadores moleculares.

- A coleção nuclear proposta para *Capsicum* spp., embora com apenas 17% dos acessos, garante índice de similaridade superior a 95%, com a coleção base, bem como diversidade fenotípica superior a esta coleção. Isso vem atender à uma expectativa presente no estabelecimento de coleções nucleares para recursos genéticos vegetais, que é o aumento das variâncias nestas coleções relativamente às coleções base.

5 CONCLUSÕES GERAIS

Este estudo demonstrou que é possível reduzir o número de descritores atualmente utilizados na caracterização dos acessos de *Capsicum* spp. da coleção de pimentas e pimentões da Embrapa Hortaliças, em cerca de 50%, sem prejuízo significativo à representação da variabilidade genética da coleção. Além disso, com não mais que 17% dos acessos, é possível constituir coleções nucleares para *Capsicum* spp., com 95% de similaridade em relação às respectivas coleções base, e até com maior diversidade fenotípica.

A análise fatorial de correspondência mostrou ser uma ferramenta alternativa apropriada para a seleção de descritores mínimos, bem como para a proposição de coleções nucleares a partir de bancos de germoplasma; sobretudo, quando estes se encontram caracterizados por dados de natureza qualitativa. Ainda assim, a proposição de coleções nucleares a partir do conhecido método de agrupamento pelo vizinho mais próximo, quase sempre, demonstra-se mais eficiente.

Em *Capsicum* spp., as variáveis relacionadas ao fruto apresentam grande variabilidade, demonstrando ser atributos muito importantes na discriminação genética de acessos deste gênero. O estudo revelou, também, que a origem geográfica do germoplasma, independentemente da espécie, tem associação direta com a divergência genética entre os acessos avaliados.

6 REFERÊNCIAS

ABADIE, T.; CORDEIRO, C. M. T.; ANDRADE, R. V. de; PARENTONI, S. N.; MAGALHES, J. R. **A coleção nuclear de germoplasma de milho para o Brasil**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2000. 37 p. (Embrapa Recursos Genéticos e Biotecnologia. Boletim de Pesquisa, 8).

ABADIE, T.; CORDEIRO, C. M. T.; FONSECA, J. R.; ALVES, R. de B. das N.; BURLE, M. L.; BRONDANI, C.; RANGEL, P. H. N.; CASTRO, E. M.; SILVA, H.T.; FREIRE, M. S.; ZIMMERMANN, F. J. P.; MAGALHÃES, J. R. Construção de uma coleção nuclear de arroz para o Brasil. **Pesquisa Agropecuária Brasileira**, Brasília, v. 40, n. 2, p. 129-136, 2005.

ALLARD, R. W. Predictive methods for germplasm identification. In: STALKER, H. T.; MURPHY, J. P. (Ed.). **Plant breeding in the 1990's**. Oxon: Wallingford, 1992. p. 119-146.

AMARAL JÚNIOR, A.T.; THIÉBAUT, J. T. L. **Análise multivariada na avaliação da diversidade em recursos genéticos vegetais**. Campos dos Goytacazes: UENF/FENORTE, 1998. 53 p.

ANDERSON, T. W. **An introduction to multivariate statistical analysis**. New York: J. Wiley & Sons, 1958. 374 p.

BASIGALUP, D. H.; BARNES, D. K.; STUCKER, R. E. Development of a core collection for perenial *Medicago* plant introductions. **Crop Science**, Madison, v. 35, n. 4, p. 1163-1168, 1995.

BEKELE, F. L.; KENNEDY, A. J.; MCDAVID, C.; LAUCKNER, F. B.; BEKELE, I. Numerical taxonomic studies on cacao (*Theobroma cacao* L.) in Trinidad. **Euphytica,** Dordrecht, v. 75, n. 3, p. 231-240. 1994.

BLANK, A. F.; SOUZA, R. J. de; GOMES, L. A. A. **Produção de pimentão em estufa**. Lavras: UFLA, 1995. 15 p. (Boletim, 55).

BORBA, T. C. O. **Diversidade genética e mapeamento associativo para produção e qualidade de grão dos acessos melhorados da coleção nuclear de arroz da Embrapa**. 2007. 109 f. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola de Agronomia e Engenharia de Alimentos, Universidade Federal de Goiás, Goiânia, 2007.

BORÉM, A. **Melhoramento de plantas.** Viçosa: Universidade Federal de Viçosa, 1997. 547 p.

BROWN, A. H. D. The case for core collections. In: BROWN, A. H. D.; FRANKEL, O. H.; MARSHALL, D. R.; WILLIAMS, J. T. (Ed.). **The use of plant genetic resources**. New York: Cambridge University Press, 1989. p.136-156.

BROWN, A. H. D. The core collection at the crossroads. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, T. J. L.; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. New York: J. Wiley, 1995. p. 3-20.

BROWN, A. H. D.; SPILLANE, C. Implementing core collections principles procedures, progress, problems and promise. In: JOHNSON, R. C.; HODGKIN, T. (Ed.). **Core collections for today and tomorrow**. Rome: IPGRI, 1999. p. 1-9.

BURTON, G. W.; DAVIES, W. E. Handling germplasm of crosspollinated forage crops. In: Holden, J. H. W.; Williams, J.T. (Ed.) **Crop genetic resources: conservation and evaluation**. George Allen & Unwin, London, 1984. p. 180-190.

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

CARVALHO, S. I. C.; BIANCHETTI, L. B.; BUSTAMANTE, P.G.; SILVA, D. B. **Catálogo de germoplasma de pimentas e pimentões (*Capsicum* spp.) da Embrapa Hortaliças**. Brasília: Embrapa Hortaliças, 2003. 49 p.

CASTINEIRAS, L. Análisis de descriptores del frijol común (*Phaseolus vulgaris*) mediante métodos multivariados. **Ciências de la Agricultura**, Havana, v. 39, n. 1, p. 54-59, 1990.

CHAIB FILHO, H. **Aplicações de análise Fatorial de Correspondências para análise de Dados**. São Carlos, SP: SBMAC, (Notas em Matemática Aplicada: 9), 2004, 60 p.

CHANDRA, S.; HUAMÁN, Z.; KRISHNA, S. H.; ORTIZ, R. Optimal sampling strategy and core collection size of Andean tetraploid potato based on isozyme data - a simulation study. **Theoretical and Applied Genetics**, Berlin, v. 104, n. 8, p. 1325-1334, 2002.

CHATFIELD, C.; COLLINS, A. J. **Introduction to multivariate analysis**. 3: ed. London: Science paperbacks, Chapman and Hall, 1986. 246 p.

CORDEIRO, C. M. T.; ABADIE, T. Coleções Nucleares. In: Nass, L. L. (Ed.). **Recursos Genéticos Vegetais**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2007. v. 1, cap. 17, p. 575-604.

CROSSA, J.; FRANCO, J. Statistical methods for classifying genotypes. **Euphytica**, Dordrecht, v. 137, n. 1, p. 19-37, 2004.

CROSSA, J.; TABA, S.; EBERHART, S. A.; BRETTING, P.; VENCOVSKY, R. Pratical consideration for maintaining germplasm in maize. **Theoretical and Applied Genetics**, Berlin, v. 89, n. 1, p. 89-95, 1994.

CROSSA, J.; DELACY, I. H.; TABA, S. The use of multivariate methods in developing a core collection. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, T. J. L.; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**, New York: J. Wiley, 1995. p. 77-92.

CRUZ, C. D. **Aplicação de algumas técnicas multivariadas no melhoramento de plantas**. 1990. 188 f. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 1990.

CRUZ, C. D. Programa Genes. **Software**. Versão Windows: aplicativo computacional em genética e estatística. Viçosa, Universidade Federal de Viçosa, 2001. 1 CD.

CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. 2: ed. Viçosa: Ed. UFV, 1994. 390 p.

CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: Ed. UFV, 2003. v. 2, 585 p.

CURY, R. **Dinâmica evolutiva e caracterização de germoplasma de mandioca (*Manihot esculenta* Crantz) na agricultura autóctone do sul do Estado de São Paulo**. Piracicaba, SP: ESALQ, 1993. 103 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 1993.

DAHER, R. F.; MORAES, C. F.; CRUZ, C. D. Seleção de caracteres morfológicos em capim-elefante (*Pennisetum purpureum Schum*). **Revista Brasileira de Zootecnia**, Viçosa, v. 26, n. 2, p. 247-259, 1997.

DAS, P. K.; GUPTA, T. D. Multivariate analysis in black grain (*Vigna mungo* (L.) Herpper). **Indian Journal of Genetics and Plant Breeding**, New Delhi, v. 44, n. 7, p. 243-247, 1984.

DIAS, L. A. S. **Divergência genética e fenética multivariada na predição de híbrido e preservação de germoplasma de cacau (*Theobroma cacao* L.)**. 1994. 94 f. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura Luiz de Queiroz – Universidade de São Paulo, Piracicaba, 1994.

DUARTE, J. B. **Princípios e utilização de técnicas multivariadas no melhoramento de plantas**. Piracicaba, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo. Monografia. 1997, 69 p.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (EMBRAPA). **Número de Bancos de Germoplasma**. Brasília: Embrapa/Cenargen, 2006. Disponível em: <http://www.cenargen.embrapa.br/recgen/curadoria/bancos.html> Acesso em: 29 nov. 2006.

ERSKINE, W.; MUEHLBAUER, F. J. Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. **Theoretical and Applied Genetics**, Berlin, v. 83, n.1, p. 119-125, 1991.

ESCOFIER, B.; PAGÈS, J. **Análisis factoriales simples y múltiples: objetivos, métodos e interpretación**. Bilbao: Servicio Editorial de la Universidad del País Vasco. 1992, 285p.

ESHBAUGH, W. H. The taxonomy of the genus *Capsicum* (Solanaceae). **Phytologia**, Austin, v. 47, n. 3, p. 153-166, 1980.

FALK, D. A. Integrated strategies for conserving plant genetic diversity. **Annals of the Missouri Botanical Garden**, Saint Louis, v. 77, p. 38-47, 1990.

FAO. Relatório da quarta reunião da Comissão de Recursos Fitogenéticos. FAO, Roma, 1991, 96 p.

FARIA, L. C. de. **Análise multivariada no estudo da divergência genética entre populações de feijão (*Phaseolus vulgaris* L.) do grupo mulatinho**. 1994. 80 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 1994.

FERREIRA, D. F. **Métodos de avaliação da divergência genética em milho e suas relações com os cruzamentos dialélicos**. 1993. 72 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura de Lavras, Lavras, 1993.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Brasília: EMBRAPA-CENARGEM, 1996, 220 p.

FILGUEIRA, F. A. R. **Novo Manual de Olericultura: agrotecnologia moderna na produção e comercialização de hortaliças**. Viçosa, 2000, 402 p.

FISHER, R. A. The precision of discriminant functions. **Annals of Eugenics**, Cambridge, v. 10, p. 422-429, 1940.

FRANKEL, O. H.; BENNETT, E. **Genetic resources in plants: their exploration and conservation**. Oxford: Blackwell (International Biological Programme, Handbock, 11), 1970, 554 p.

FRANKEL, O. H.; SOULE, M. E. **Conservation and evolution**. Londres: Cambridge University Press, 1981, 327 p.

FRANKEL, O. H. Genetic perspectives of germplasm conservation. In: ARBER, W. K.; LLIMENSEE, K.; PEACOCK, W. J.; STARLINGER, P. (Eds.). **Genetic manipulation: impact on man and society**. Cambridge: Cambridge University, 1984. p. 161-170.

FRANKEL, O. H.; BROWN, A. H. D. Plant genetic resources today: a critical reappraisal. In: HOLDEN, J. H. W.; WILLIAMS, J. T. (Ed.). **Crop genetic resources: conservation and evaluation**. London: G. Allen & Unwin, 1984, p. 249-257.

FRANKEL, O. H.; BROWN, A. H. D.; BURDON, J. J. **The conservation of plant biodiversity**. Cambrigde: Cambrigde University, 1995, 299p.

GALWEY, N. W. Verifying and validating the representativeness of a core collection. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, T. J. L.; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. New York: J. Wiley, p. 187-198, 1995.

GOWER, J. C. 1971. A general coefficient of similarity and some of its properties. **Biometrics**. Washington, v. 27, n. 4, p. 857–874.

HARLAN, J.; de WET, J. M. J. A simplified classification of cultivated Sorghum. **Crop Science**, Madison, v. 12, p. 172-176, 1972.

HARCH, B. D.; BASFORD, K. E. DELACY, I. H.; LAWRENCE, P. K.; CRUICKSHANK, A. Mixed data types and the use of pattern analysis on the Australian groundnut germplasm data. **Genetic Resources and Crop Evolution**, Dordrecht, v. 43, n, 4, p. 363-376, 1996.

HAWKES, J. G. Germplasm collection, preservetion and use. In FREI, K. J. (Ed.). **Plant Breeding II**. Kaliany Publishers, Ludhiana, 1982. p. 57-83.

HIRSCHFELD, H. O. A connection between correlation and contingency. **Proceedings of the Cambridge Philosophical Society, Mathematical and Physical Sciences**, Cambridge, v. 31, p. 520-524, 1935.

HOLCOMB, J.; TOLBERT, D. M.; JAIN, S. K. A diversity analysis of genetic resources in rice. Euphytica, Dordrecht, v. 26, n. 2, p. 441-450, 1977.

HOTELLING, H. Analysis of a complex of statistical variables in principal components. **Journal of Education Psychology**, Washington, v. 24, n. 6, p. 417-441, 1933.

HOTELLING, H. Simplified calculation of principal components. **Psychometrika**, Greensboro, v. 1, n. 1, p. 27-35, 1936.

HOWES, C. Guidelines for developing descriptors lists. **Plant Genetic Resources Newsletter**, Roma, n. 45. p. 26-32, 1981.

HU, J.; ZHU, J.; XU, H. M. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. **Theoretical and Applied Genetics**, Berlin, v. 101, n. 1-2, p. 264-268, 2000.

IBPGR. **Descriptors for Phaseolus vulgaris**. Rome, International Board for Plant Genetic Resources, 1982, 45 p.

IBPGR. **Annual Report 1986**. Rome, International Board for Plant Genetic Resources, 1987.

IPGRI. **Descriptores para Capsicum (*Capsicum* spp.)**. Roma: International Board for Plant Genetic Resources, 1995, 51 p.

JACKSON, M. T.; PHAM, J. L.; NEWBURY, H. J.; FORD-LLOYD, B.V.; VIRK, P.S. A core collection for rice-needs opportunities and constraints. In: JOHNSON, R. C.; HODGKIN, T. (Ed.) **Core collections for today and tomorrow**. Roma: IPGRI, 1999. p. 18-27.

JANA, S.; SINGH, K. B. Evidence of geographical divergence in Kabuli chickpea from germplasm evaluation data. **Crop Science**, Madison, v. 33, p. 626-632, 1993.

JOLLIFFE, I. T. Discarding variables in a principal component analysis, I Artificial data. **Applied Statistics**, London, v. 21, p. 160-173, 1972.

JOLLIFFE, I. T. Discarding variables in a principal component analysis, II Real data. **Applied Statistics**, London, v. 22, p. 21-31, 1973.

JOHNSON, R. A. & WICKERN, D. W. **Applied multivariate statistical analysis**. Englewood Cliffs, Prentice Hall, 1988. 607 p.

JUDEZ, A. L. **Técnica de análisis de datos multidimensionales**. Ministerio de agricultura, pesca y alimentación. Madrid, 1989, 301 p.

KRESOVICH, S.; McFERSON, J. R. Assessment and management of plant genetic diversity: considerations of intra and interspecific variation. **Field Crops Research**, Amsterdam, v. 29, n. 3, p. 185-204, 1992.

LI, Y.; WU, S.; CAO, Y. Cluster analysis of an international collection of foxtail millet (Setaria italica (L.) P. Beauv.). **Euphytica**, Dordrecht, v. 83, n. 1, p. 79-85, 1995.

LIU, F.; BOTHMER, R. von; SALOMON, B. Genetic diversity among East Asian accessions of the barley core collection as revealed by six isozyme loci. **Theoretical and Applied Genetics**, Berlin, v. 98, n. 8, p. 1226-1233, 1999.

LUZ, F. J. F. **Caracterizações morfológica e molecular de pimentas (Capsicum chinense Jack.)**. 2007. 150 f. Tese (Doutorado em Produção Vegetal) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista Júlio de Mesquita Filho, Jaboticabal, 2007.

MALOSETTI, M.; ABADIE, T. Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphological traits. **Genetic Resources and Crop Evolution**, Dordrecht, v. 48, n. 4, p. 381-390, 2001.

MANLY, B. F. J. **Randomization, bootstrap and Monte Carlo methods in biology**. London: Chapman & Hall, 1997. 281 p.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. **Agronegócio brasileiro: uma oportunidade de investimentos**. Brasília: MAPA, 2004. Disponível em: <http://www.agricultura.gov.br/portal/page?_pageid=33,968707&_dad=portal&_schema=PORTAL>. Acesso em: 21 agosto 2006.

MARCHIORO, V. S.; CARVALHO, F. I. F. de; OLIVEIRA, A. C.; CRUZ, P. J.; LORENCETTI, C.; BENIN, G.; SILVA, J. A. G.; SCHMIDT, D. A. M. Dissimilaridade genética entre genótipos de aveia. **Ciência e Agrotecnologia**, Lavras, v. 27, n. 2, p. 285-294, 2003.

MARTINELLO, G. E. ; LEAL, N. R.; AMARAL JR, A. T.; PEREIRA, M. G.; DAHER, R. F. Divergência genética em acessos de quiabeiro com base em marcadores morfológicos. **Horticultura Brasileira**, Brasília, v. 20, n. 1, p. 52-58, 2002.

McNEELY, J. A.; MILLER, K. R.; REID, W. V.; MITTERMEIER, R. A.; WERNER, T. B. **Conserving world's biological diversity**. Gland: International Union for Conservation of Nature and Natural Resource, 1990, 193 p.

NASS, L. L.; PELLICANO, I. J.; VALOIS, A. C. C. Utilization of genetic resources for maize and soybean breeding in Brazil. **Revista Brasileira de Genética**, Ribeirão Preto, v. 16, n. 4, p. 983-988, 1993.

OLIVEIRA, E. J. **Análise multivariada no estudo da divergência genética entre cultivares de feijão (*Phaseolus vulgaris* L.)**. 1989. 91 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Universidade Federal de Viçosa, Viçosa. 1989.

OLIVEIRA, M. F. **Avaliação de cinco estratégias de amostragem para obtenção da coleção nuclear de soja (*Glycine max* (L.) Merrill)**. 2007. 143 f. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura Luiz de Queiroz – Universidade de São Paulo, Piracicaba, 2007.

ONYILAGHA, J. C. Numerical analysis of variation among Nigerian *Dioscorea rotunda* accessions. **Euphytica**, Dordrecht, v. 35, n. 2, p. 413-419, 1986.

PEETERS, J. P.; WILLIAMS, J. T. Toward better use of genebanks with special reference to information. **Plant Genetic Resources Newsletter**, Roma, v. 60, p. 22-32, 1984.

PEREIRA, A. V. **Utilização de análise multivariada na caracterização de germoplasma de mandioca (*Manihot esculenta* Crantz.)**. 1989. 180 f. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 1989.

PEREIRA, A. V.; VENCOVSKY, R.; CRUZ, C. D. Selection of botanical and agronomical descriptors for the characterization of cassava (*Manihot esculenta* Crantz.) germplasm. **Revista Brasileira de Genética**, Ribeirão Preto, v. 15, n. 1, p. 115-124, 1992.

PERRY, M. C.; McINTOSH, M. S. Geographical patterns of variation in the USDA soybean germplasm collection: I. Morphological traits. **Crop Science**, Madison, v. 31, n. 5, p. 1350-1355, 1991.

QUEIROZ, M. A.; LOPES, M. A. Importância dos recursos genéticos vegetais para o agronegócio. In: Nass, L. L. (Ed.). **Recursos Genéticos Vegetais**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2007. v. 1, cap. 2, p. 61-119.

RABBANI, M. A.; IWABUCHI, A.; MURAKAMI, Y.; SUZUKI, T.; TAKAYANAGI, K. Phenotypic variation and the relationships among mustard (*Brassica juncea* L.) germplasm from Pakistan. **Euphytica**, Dordrecht, v. 101, n. 3, p. 357-366, 1998.

RAMOS, S. R. R. ; QUEIROZ, M. A. Caracterização morfológica: experiência do BAG de cucurbitáceas da Embrapa Semi - Árido, com acessos de abóbora e moranga. **Horticultura Brasileira**, Brasília, v. 17, suplemento, p. 9-12, 1999.

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

RAO, A. V.; PRASSAD, A. S. R.; SAI KRISHNA, T.; SESHU, D. V.; SRINIVASAN, T. E. Genetic divergence among some brown planthopper resistant rice varieties. **Indian Journal of Genetics and Plant Breeding**, New Delhi, v.41, n.2, p.179-185, 1981.

RAO, K. E. P.; MENGESHA, M. H.; REDDY, V. G. International use of a sorghum germplasm collection. In: BROWN, A. H. D.; FRANKEL, O. H.; MARSHALL, D. R.; WILLIAMS, J. T. (Ed.). **The use of plant genetic resources**. New York: Cambridge University Press, 1989, p. 49-67.

REIFSCHNEIDER, F. J. B. (Org). ***Capsicum*: pimentas e pimentões no Brasil**. Brasília: Embrapa Comunicação para Transferência de Tecnologia: Embrapa Hortaliças, 113 p. 2000.

RESENDE, M. D. V.; DUARTE , J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical,** Goiânia, v. 37, n. 3, p. 182-194, 2007.

RIBEIRO, R. M. A. Glossário de termos de coleta e conservação de recursos genéticos. **Ciência da Informação (Comunicações)**, Brasília, v. 24, n. 3, 1995. p. 1-7.

RODRIGUES, D. P.; ASTOLFI FILHO, S.; CLEMENT, C. R. Molecular marker-mediated validation of morphologically defined landraces of Pejibaye (*Bactris gasipaes*) and their phylogenetic relationships. **Genetic Resources and Crop Evolution**, Dordrecht, v. 51, n. 8, p. 871-882, 2005.

ROHLF, F. J. NTSYSpc. **Software**. Version 2.02g. 1989.

SAS INSTITUTE. SAS® System for Windows, versão 8.0: **software**. SAS Institute Inc., 2002. Conjunto de programas. 2 CD-ROM.

SILVA, M. L.; QUEIRÓZ, M. A.; FERREIRA, M. A. J. F.; BUSO, G. S. C. Caracterização morfológica e molecular de acessos de melancia. **Horticultura Brasileira**, Brasília, v. 24, n. 4, p. 405-409, 2006.

SIMPSON, C. E.; HOWARD, E. R.; HIGGINS, D. L.; THOMAS, G. D. **Catalog of minimum descriptors of *Arachis hypogaea* L. germplasm collections from South America**. College Station: Texas A&M University, p. 1977-1982, 1986.

SINGH, S. P.; GUTIERREZ, J. A.; MOLINA, A.; URREA, C.; GEPTS, P. Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. **Crop Science**, Madison, v. 31, p. 23-29, 1991.

SLOTEN, D. H. van. **The role of curators, breeders and other users of germplasm in characterization and evaluation of crop genetic resources**. Rome: IBPGR/SEAN, 1987. p. 3-8.

SMITH, J. S. C. Genetic variability within U.S. hybrid maize: Multivariate analysis of isozyme data. **Crop Science**, Madison, v. 24, p. 1041–1046, 1984.

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

SOUZA, J.; PENALOZA, R. A. S. **Estatística exploratória de séries nominais e ordinais: teoria de aplicação a medidas multidimensionais nebulosas de pobreza para dados ordinais**. Disponível em: www.ufrgs.br/ppge/pcientifica/2005_03.pdf Acesso em: 11 de fevereiro de 2008.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics: a biometrical approach**. New York: McGraw-Hill, 1980, 634 p.

STRAPASSON, E.; VENCOVSKY, R.; BATISTA, L. A. R. Seleção de Descritores na Caracterização de Germoplasma de Paspalum sp. por meio de Componentes Principais. **Revista Brasileira de Zootecnia**, Viçosa, v. 29, n. 2, p. 373-381, 2000.

SUDRÉ, C. P. **Divergência genética e avaliação da resistência à mancha bacteriana em *Capsicum* spp**. 2003. 112 f. Tese (Doutorado em Produção Vegetal) – Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, 2003.

SUDRÉ, C. P.; CRUZ, C. D.; RODRIGUES, R.; RIVAL, E. M.; AMARAL JÚNIOR, A. T. do; SILVA, D. J. H.; PEREIRA, T. N. S. Variáveis multicategóricas na determinação da divergência genética entre acessos de pimenta e pimentão. **Horticultura Brasileira**, Brasília, v. 24, n. 1, p. 88-93, 2006.

TEIXEIRA, R. **Diversidade em *Capsicum*: análise molecular, morfoagronômica e química**. 1996. 81 f. Tese (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, 1996.

THOMPSON, S. K. **Sampling**. 2: ed. New York: Wiley, 2002. 367p.

UPADHYAYA, H. D.; ORTIZ, R. A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. **Theoretical and Applied Genetics**, Berlin, v. 102, n. 8, p. 1292-1298, 2001.

VALLS, J. F. M. Caracterização morfológica, reprodutiva e bioquímica de germoplasma vegetal. In: ENCONTRO SOBRE RECURSOS GENÉTICOS, 1, 1988, Jaboticabal. **Anais...** Jaboticabal: FCAV, 1988. p. 106-124.

VALLS, J. F. M. Caracterização de Recursos Genéticos Vegetais. In: Nass, L. L. (Ed.). **Recursos Genéticos Vegetais**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2007. v. 1, cap. 8, p. 281-305.

VAN HINTUM, T. J. L.; BROWN, A. H. D.; SPILLANE, C.; HODGKIN, T. **Core collections of plant genetic resources**. (IPGRI. Technical bulletin, 3), Rome: International Plant Genetic Resources Institute, 2000. 48p.

VAN SLOTEN, D. H. IBPGR/SEAN. **Special issue**, p. 3-8, 1987.

VASCONCELOS, A. M. N. **Análise de Correspondência simples e múltipla**. Brasília, Departamento de Estatística, Universidade de Brasília – Apostila. 2000. 30 p.

VASCONCELOS, E. S. de; CRUZ, C. D.; BHERING, L. L.; FERREIRA, A. Estratégias de amostragem e estabelecimento de coleções nucleares. **Pesquisa Agropecuária Brasileira**, Brasília, v. 42, n. 4, p. 507-514, 2007.

VILELA-MORALES, E. A. Documentação e informática de recursos genéticos. In: ENCONTRO SOBRE RECURSOS GENÉTICOS, 1., 1988, **Anais**... Jaboticabal: UNESP-FCAV, 1988. p. 135-147.

VILELA-MORALES, E. A.; VALOIS, A. C. C.; COSTA, I. R. S. **Regras para denominação e codificação de germoplasma vegetal**. Brasília: EMBRAPA-CENARGEN, v. 1, 1993.

VILELA-MORALES, E. A.; VALOIS, A. C. C. Recursos genéticos vegetais autóctones e seus usos no desenvolvimento sustentável. Brasília: **Cadernos de Ciência & Tecnologia**, v. 17, n. 2, p. 11-42, 2000.

VON DER HEIZDEN, P. G. M.; DE FALGUEROLLES, A.; DE LEEUW, J. A combined approach to contingency table analysis using correspondence analysis (with Discussion). **Applied Statistics**, London, v. 38, n. 1, p. 249-292, 1989.

WANG, J. C.; HU, J.; XU, H. M.; ZHANG, S. A strategy on constructing core collections by least distance stepwise sampling. **Theoretical and Applied Genetics**. Berlin, v. 115, n. 1, p. 1-8, 2007.

WILLIAMS, J. A decade of crop genetic resources research. In: **Crop Genetic Resources: Conservation & Evaluation**, Holden &Williams (eds). London: George Allen & Unwin London, 1984.

WISHART, D. Hierarchical cluster analysis with messy data. In Gaul, W.; Schader, M. (Ed.): **Classification as a Tool of Research**. New York: Elsevier Science Publishers. p. 453-460, 1986.

XU, H. M.; QIU, Y. X.; HU, J.; WANG, J. C. Methods of Constructing Core Collection of Crop Germplasm by Comparing Different Genetic Distances, Cluster Methods and Sampling Strategies. **Acta Agronomica Sinica**, Beijing, v. 30, n. 9, p. 932-936, 2004.

YONEZAWA, K.; NOMURA, T.; MORISHIMA, H. Sampling strategies for use in stratified germplasm collections. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, T. J. L.; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. New York: J. Wiley, 1995. p. 35-54.

ZEULI, P. L. S.; QUALSET, C. O. The durum wheat core collection and the plant breeder. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, T. J. L.; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. New York: J. Wiley, 1995. p. 213-228.

APÊNDICES

Apêndice A. Descritores morfológicos aplicados na caracterização da coleção de germoplasma de *Capsicum* spp. da Embrapa Hortaliças e suas respectivas modalidades ou categorias

Descritor	valor
1- Procedência	
Ásia	1
África	2
Europa	3
EUA e Canadá	4
México	5
América Central	6
Am. do Sul (exc. Brasil)	7
Brasil (Norte)	8
Brasil (Nordeste)	9
Brasil (Centro-Oeste)	10
Brasil (Sudeste)	11
Brasil (Sul)	12
2- Espécie	
C. annuum	1
C. baccatum	2
C. chinense	3
C. frutescens	4
3-Cor da Haste	
verde	1
verde c/ estrias violetas	2
violeta	3
4- Antocianina nodal	
verde	1
violeta claro	3
violeta	5
violeta escuro	7
5-Formato da haste	
cilíndrica	1
angulada	2
alado	3
6-Pubescência da haste	
esparsa	3
intermediária	5
densa	7
7-Comp. planta (cm)	
abaixo de 25	1
25 - 45	2
46 - 65	3
66 - 85	4
acima de 85	5
8-Hábito crescimento	
prostrado	3
intermediário	5
ereto	7
outro	9
9- Comp. dossel (cm)	
até 80	1
>80 a 120	2
>120 a 160	3
acima de 160	4
10-Comp. haste (cm)	
até 15	1
>15 A 30	2
>30 a 50	3
acima de 50	4
11-Diâm. haste (cm)	
até 1	1
>1 a 2	2
>2 a 3	3
acima de 3	4
12-Densid. ramificação	
esparsa	3
intermediária	5
densa	7
13-Brotação abaixo da primeira bifurcação	
ausente	1
esparsa	3
intermediária	5
densa	7
14-Densidade de folha	
esparsa	3
intermediária	5
densa	7
15-Cor da folha	
amarelo	1
vede claro	2
verde	3
verde escuro	4
violeta-claro	5
violeta	6
mesclado	7
verde c/ antocianina	8
16-Forma da folha	
deltóide	1
ovalada	2
lanceolada	3
17-Pubescência folha	
esparsa	3
intermediária	5
densa	7
18- Dias início do florescimento	
até 30	1
31 a 60	2
61 a 90	3
acima de 90	4
19-Nº de flores/axila	
uma	1
duas	2
três ou mais	3
muitas c/ entre-nó curto	4
uma e duas	5
uma, duas e três	6
duas e três	7
duas, três e quatro	8
20-Pose da flor	
pendente	3
intermediária	5
ereta	7
todas	8
intermediária e ereta	9
pendente e intermediária	10
pendente e ereta	11
21-Cor da corola	
branco	1
amarelo claro	2
amarelo	3
amarelo esverdeado	4
violeta c/ base branca	5
branco c/ base violeta	6
branco c/ margem violeta	7
violeta	8
branco esverdeado	9
branco c/ mancha purpura	10
branco esverdeado c/ mancha purpura	11
22-Cor mancha da corola	
branco	1
amarelo	2
verde amarelado	3
verde	4
violeta	5
sem mancha	6
23-Forma da corola	
rotada	1
campanulada	2
intermediária	3
24-Cor da antera	
branco	1
amarelo	2
azul pálido	3
azul	4
violeta	5
amarelo c/ mancha azul	6
claro	6
25-Cor do filamento	
branco	1
amarelo	2
verde	3
azul	4
violeta claro	5
violeta	6
azul violeta	7
26-Posição do estigma	
inserto	3
mesmo nível	5
excerto	7
mesmo nível e excerto	8
inserto e mesmo nível	9
inserto e excerto	10
27-Pigmento do cálice	
ausente	0
presente	1
28-Margem do cálice	
inteiro	1
intermediário	2
dentado	3
29- Constrição anelar no cálice	
ausente	0
presente	1
30-Dias início da frutificação	
até 60	1
61 a 90	2
91 a 120	3
acima de 120	4
31- Cor fruto imaturo	
branco	1
amarelo	2
verde	3
laranja	4
violeta	5
violeta escuro	6
amarelo esverdeado	7
verde amarelado	8
branco amarelado	9
marrom	10
32-Posição do fruto	
pendente	3
intermediário	5
ereta	7
todas	8
pendente e intermediária	9
pendente e ereto	10
intermediário e ereto	11
33-Cor fruto maduro	
branco	1
amarelo limão	2
amarelo laranja pálido	3
amarelo laranja	4
laranja pálido	5
laranja	6
vermelho claro	7
vermelho	8
vermelho escuro	9
violeta	10
marrom	11
preto	12
amarelo	13
amarelo pálido	14
34-Formato do fruto	
alongado	1
arredondado	2
triangular	3
campanulada	4
retangular	5
tipo pitanga	6
35-Comprimento do fruto (cm)	
até 1	1
>1 a 2	2
>2 a 4	3
>4 a 8	4
>8 a 12	5
acima de 12	6
36-Largura do fruto	
até 1	1
>1 a 2,5	2
>2,5 a 5	3
>5 a 8	4
acima de 8	5
37-Peso do fruto (g)	
até 1	1
>1 a 3	2
>3 a 9	3
>9 a 27	4
>27 a 81	5
acima de 81	6
38-Comp. do pedúnculo (cm)	
até 2	1
>2 a 4	2
>4 a 6	3
acima de 6	4
39-Espessura da parede (mm)	
até 1	1
>1 a 2	2
>2 a 3	3
>3 a 4	4
>4 a5	5
acima de 5	6
40-Ombro do fruto	
agudo	1
obtuso	2
truncado	3
cordato	4
lobato	5
41- Pescoço na base do fruto	
ausente	0
presente	1
42-Formato da ponta do fruto	
pontiagudo	1
truncado (“blunt”)	2
afundado	3
afundado c/ ponta	4
43- Apêndice na ponta do fruto	
ausente	0
presente	1
44-Secção transversal	
levemente corrugado	3
intermediário	5
corrugado	7
45-Número de lóculos	
um	1
dois	2
três	3
quatro	4
cinco	5
46-Superfície do fruto	
liso	1
semi-rugoso	2
rugoso	3
liso c/ estrias	4
semi-rugoso c/ estrias	5
47-Persistência entre fruto/ pedicelo	
pouco persistente	3
intermediário	5
persistente	7
48-Comp. da placenta	
até ¼	1
de ¼ a ½	2
acima de ½	3
49-Pungência	
doce	1
picante baixo	2
picante médio	3
picante alto	4
50- Aroma	
baixo	1
médio	2
alto	3
51- Cor da semente	
amarelo	1
marrom	2
preto	3
outros	4
52-Superfície da semente	
lisa	1
rugosa	2
corrugada	3
53-Número sementes por fruto	
menor de 20	1
20 a 50	2
acima de 50	3
54-Segregação	
ausência de segregação	0
frutos c/ formatos diferentes	1
plantas c/ e sem entre-nós curtos	2
frutos c/ cores diferentes	3
fruto e planta c/ e sem antocianina	4
flores c/ cores diferentes	5
frutos c/ posições diferentes	6
55-Mistura varietal	
ausente	0
presente	1
56-Macho-esterilidade	
Ausente	0
Presente	1

Apêndice B. Rotina computacional para implementação da análise de correspondência múltipla, no aplicativo SAS (*Statistical Analysis System Institute*, 2002), relativa a dados de caracterização de acessos do Banco de Germoplasma de *Capsicum* spp. da Embrapa Hortaliças (2007).

```
data Tese;
input  Aces $ 1-5 Dfl $ 6-9 Dft $ 10-13 EFt $ 14-17 CHt $ 18-21 DHt $ 22-
25 LPt $ 26-29 NLL $ 30-33 CFt $ 34-37 LFt $ 38-41 PFt $ 42-45 CPF $ 46-
49 Seg $ 50-53 Pro $ 54-57 Esp $ 58-61 CrH $ 62-65 AtN $ 66-69 FHt $ 70-
73 PHt $ 74-77 CPt $ 78-81 Hbt $ 82-85 DsR $ 86-89 BtF $ 90-93 DsF $ 94-
97 CrF $ 98-101 FmF $ 102-105 PbF $ 106-109 NFA $ 110-113 PFr $ 114-117
CrC $ 118-121 CCR $ 122-125 FmC $ 126-129 Cat $ 130-133 CFn $ 134-137 Pet
$ 138-141 McE $ 142-145 Pgl $ 146-149 MgC $ 150-153 CCL $ 154-157 CFI $
158-161 PsF $ 162-165 APF $ 166-169 CFM $ 170-173 FFt $ 174-177 Oft $
178-181 PCF $ 182-185 FPF $ 186-189 STv $ 190-193 SFt $ 194-197 PtF $
198-201 CPl $ 202-205 MVr $ 206-209 Pug $ 210-213 Arm $ 214-217 CSt $
218-221 SSt $ 222-225 NSF $ 226-229;
array v {267} Dfa Dfb Dfc Dfd Fta Ftb Ftc Ftd Epa Epb Epc Epd Epe Epf CHa
CHb CHc CHd DHa DHb DHc DHd LPa LPb LPc LPd NLa NLb NLc NLd NLe CFa CFb
CFc CFd CFe CFF LFa LFb LFc LFd LFe PFa PFb PFc PFd PFe PFf CPa CPb CPc
CPd Sga Sgb Sgc Sgd Sge Sgf Sgg Pra Prb Prc Prd Pre Prf Prg Prh Pri Prj
Prk Prl Epa Epb Epc Epd CHa CHb CHc Ata Atb Atc Atd FHa FHb FHc PHa PHb
PHc CPa CPb CPc CPd CPe Hba Hbb Hbc Hbd Dsa Dsb Dsc Bta Btb Btc Btd DFa
DFb DFc CFa CFb CFc CFd CFe CFf CFg CFh FFa FFb FFc PFa PFb PFc NXa NXb
NXc NXd NXe NXf NXg NXh PFa PFb PFc PFd PFe PFf PFg CCa CCb CCc CCd CCe
CCf CCg CCh CCi CCj CCk CMa CMb CMc CMd CMe CMf FCa FCb FCc CAa CAb CAc
CAd CAe CAf CFa CFb CFc CFd CFe CFf CFg PEa PEb PEc PEd PEe PEf MEa MEb
PCa PCb MCa MCb MCc CAa CAb CIa CIb CIc CId CIe CIf CIg CIh CIi CIj PFa
PFb PFc PFd PFe PFf PFg APa APb CMa CMb CMc CMd CMe CMf CMg CMh CMi CMj
CMk CMl CMm CMn FFa FFb FFc FFd FFe FFf OFa OFb OFc OFd OFe PBa PBb FPa
FPb FPc FPd STa STb STc SFa SFb SFc SFd SFe PFa PFb PFc CPa CPb CPc MVa
MVb Pga Pgb Pgc Pgd Ara Arb Arc CSa CSb CSc CSd SSa SSb SSc NSa NSb NSc
;
do i = 1 to 267;
 v (i) = 0;
 end;
 if      Dfl = 'Dfa' then Dfa = 1;
 else if Dfl = 'Dfb' then Dfb = 1;
 else if Dfl = 'Dfc' then Dfc = 1;
 else if Dfl = 'Dfd' then Dfc = 1;
 else if Dfl = '   ' then do;
 Dfa = 1/4;
 Dfb = 1/4;
 Dfc = 1/4;
 Dfd = 1/4;
 end;
 if      Dft = 'Fta' then Fta = 1;
 else if Dft = 'Ftb' then Ftb = 1;
 else if Dft = 'Ftc' then Ftc = 1;
 else if Dft = 'Ftd' then Ftd = 1;
 else if Dft = '   ' then do;
 Fta = 1/4;
 Ftb = 1/4;
 Ftc = 1/4;
 Ftd = 1/4;
 end;
 if      EFt = 'Epa' then Epa = 1;
 else if EFt = 'Epb' then Epb = 1;
 else if EFt = 'Epc' then Epc = 1;
```

```
 else if EFt = 'Epd' then Epd = 1;
 else if EFt = 'Epe' then Epe = 1;
 else if EFt = 'Epf' then Epf = 1;
 else if EFt = '   ' then do;
 Epa = 1/6;
 Epb = 1/6;
 Epc = 1/6;
 Epd = 1/6;
 Epe = 1/6;
 Epf = 1/6;
end;
 if      CHt = 'CHa' then CHa = 1;
 else if CHt = 'CHb' then CHb = 1;
 else if CHt = 'CHc' then CHc = 1;
 else if CHt = 'CHd' then CHd = 1;
 else if CHt = '   ' then do;
 CHa = 1/4;
 CHb = 1/4;
 CHc = 1/4;
 CHd = 1/4;
end;
 if      DHt = 'DHa' then DHa = 1;
 else if DHt = 'DHb' then DHb = 1;
 else if DHt = 'DHc' then DHc = 1;
 else if DHt = 'DHd' then DHd = 1;
 else if DHt = '   ' then do;
 DHa = 1/4;
 DHb = 1/4;
 DHc = 1/4;
 DHd = 1/4;
end;
 if      LPt = 'LPa' then LPa = 1;
 else if LPt = 'LPb' then LPb = 1;
 else if LPt = 'LPc' then LPc= 1;
 else if LPt = 'LPd' then LPd = 1;
 else if LPt = '   ' then do;
 LPa = 1/4;
 LPb = 1/4;
 LPc = 1/4;
 LPd = 1/4;
end;
 if      NLL = 'NLa' then NLa = 1;
 else if NLL = 'NLb' then NLb = 1;
 else if NLL = 'NLc' then NLc = 1;
 else if NLL = 'NLd' then NLd = 1;
 else if NLL = 'NLe' then NLe = 1;
 else if NLL = '   ' then do;
 NLa = 1/5;
 NLb = 1/5;
 NLc = 1/5;
 NLd = 1/5;
 NLe = 1/5;
end;
 if      CFt = 'CFa' then CFa = 1;
 else if CFt = 'CFb' then CFb = 1;
 else if CFt = 'CFc' then CFc = 1;
 else if CFt = 'CFd' then CFd = 1;
 else if CFt = 'CFe' then CFe = 1;
 else if CFt = 'CFf' then CFf = 1;
 else if CFt = '   ' then do;
```

```
 CFa = 1/6;
 CFb = 1/6;
 CFc = 1/6;
 CFd = 1/6;
 CFe = 1/6;
 CFf = 1/6;
end;
 if      LFt = 'LFa' then LFa = 1;
 else if LFt = 'LFb' then LFb = 1;
 else if LFt = 'LFc' then LFc = 1;
 else if LFt = 'LFd' then LFd = 1;
 else if LFt = 'LFe' then LFe = 1;
 else if LFt = '   ' then do;
 LFa = 1/5;
 LFb = 1/5;
 LFc = 1/5;
 LFd = 1/5;
 LFe = 1/5;
end;
 if      PFt = 'PFa' then PFa = 1;
 else if PFt = 'PFb' then PFb = 1;
 else if PFt = 'PFc' then PFc = 1;
 else if PFt = 'PFd' then PFd = 1;
 else if PFt = 'PFe' then PFe = 1;
 else if PFt = 'PFf' then PFf = 1;
 else if PFt = '   ' then do;
 PFa = 1/6;
 PFb = 1/6;
 PFc = 1/6;
 PFd = 1/6;
 PFe = 1/6;
 PFf = 1/6;
end;
 if      CPF = 'CPa' then CPa = 1;
 else if CPF = 'CPb' then CPb = 1;
 else if CPF = 'CPc' then CPc = 1;
 else if CPF = 'CPd' then CPd = 1;
 else if CPF = '   ' then do;
 CPa = 1/4;
 CPb = 1/4;
 CPc = 1/4;
 CPd = 1/4;
end;
 if      Seg = 'Sga' then Sga = 1;
 else if Seg = 'Sgb' then Sgb = 1;
 else if Seg = 'Sgc' then Sgc = 1;
 else if Seg = 'Sgd' then Sgd = 1;
 else if Seg = 'Sge' then Sge = 1;
 else if Seg = 'Sgf' then Sgf = 1;
 else if Seg = 'Sgg' then Sgg = 1;
 else if Seg = '   ' then do;
  Sga = 1/7;
  Sgb = 1/7;
  Sgc = 1/7;
  Sgd = 1/7;
  Sge = 1/7;
  Sgf = 1/7;
  Sgg = 1/7;
end;
```

```
 if      Pro = 'Pra' then Pra = 1;
 else if Pro = 'Prb' then Prb = 1;
 else if Pro = 'Prc' then Prc = 1;
 else if Pro = 'Prd' then Prd = 1;
 else if Pro = 'Pre' then Pre = 1;
 else if Pro = 'Prf' then Prf = 1;
 else if Pro = 'Prg' then Prg = 1;
 else if Pro = 'Prh' then Prh = 1;
 else if Pro = 'Pri' then Pri = 1;
 else if Pro = 'Prj' then Prj = 1;
 else if Pro = 'Prk' then Prk = 1;
 else if Pro = 'Prl' then Prl = 1;
 else if Pro = '   ' then do;
 Pra = 1/12;
 Prb = 1/12;
 Prc = 1/12;
 Prd = 1/12;
 Pre = 1/12;
 Prf = 1/12;
 Prg = 1/12;
 Prh = 1/12;
 Pri = 1/12;
 Prj = 1/12;
 Prk = 1/12;
 Prl = 1/12;
end;
 if      Esp = 'Epa' then Epa = 1;
 else if Esp = 'Epb' then Epb = 1;
 else if Esp = 'Epc' then Epc = 1;
 else if Esp = 'Epd' then Epd = 1;
 else if Esp = '   ' then do;
 Epa = 1/4;
 Epb = 1/4;
 Epc = 1/4;
 Epd = 1/4;
end;
 if      CrH = 'CHa' then CHa = 1;
 else if CrH = 'CHb' then CHb = 1;
 else if CrH = 'CHc' then CHc = 1;
 else if CrH = '   ' then do;
 CHa = 1/3;
 CHb = 1/3;
 CHc = 1/3;
end;
 if      AtN = 'Ata' then Ata = 1;
 else if AtN = 'Atb' then Atb = 1;
 else if AtN = 'Atc' then Atc = 1;
 else if AtN = 'Atd' then Atd = 1;
 else if AtN = '   ' then do;
 Ata = 1/4;
 Atb = 1/4;
 Atc = 1/4;
 Atd = 1/4;
end;
 if      FHt = 'FHa' then FHa = 1;
 else if FHt = 'FHb' then FHb = 1;
 else if FHt = 'FHc' then FHc = 1;
 else if FHt = '   ' then do;
 FHa = 1/3;
 FHb = 1/3;
```

```
 FHc = 1/3;
end;
 if      PHt = 'PHa' then PHa = 1;
 else if PHt = 'PHb' then PHb = 1;
 else if PHt = 'PHc' then PHc = 1;
 else if PHt = '   ' then do;
 PHa = 1/3;
 PHb = 1/3;
 PHc = 1/3;
end;
 if      CPt = 'CPa' then CPa = 1;
 else if CPt = 'CPb' then CPb = 1;
 else if CPt = 'CPc' then CPc = 1;
 else if CPt = 'CPd' then CPd = 1;
 else if CPt = 'CPe' then CPe = 1;
 else if CPt = '   ' then do;
 CPa = 1/5;
 CPb = 1/5;
 CPc = 1/5;
 CPd = 1/5;
 CPe = 1/5;
end;
 if      Hbt = 'Hba' then Hba = 1;
 else if Hbt = 'Hbb' then Hbb = 1;
 else if Hbt = 'Hbc' then Hbc = 1;
 else if Hbt = 'Hbd' then Hbd = 1;
 else if Hbt = '   ' then do;
 Hba = 1/4;
 Hbb = 1/4;
 Hbc = 1/4;
 Hbd = 1/4;
end;
 if      DsR = 'Dsa' then Dsa = 1;
 else if DsR = 'Dsb' then Dsb = 1;
 else if DsR = 'Dsc' then Dsc = 1;
 else if DsR = '   ' then do;
 Dsa = 1/3;
 Dsb = 1/3;
 Dsc = 1/3;
end;
 if      BtF = 'Bta' then Bta = 1;
 else if BtF = 'Btb' then Btb = 1;
 else if BtF = 'Btc' then Btc = 1;
 else if BtF = 'Btd' then Btd = 1;
 else if BtF = '   ' then do;
 Bta = 1/4;
 Btb = 1/4;
 Btc = 1/4;
 Btd = 1/4;
end;
 if      DsF = 'DFa' then DFa = 1;
 else if DsF = 'DFb' then DFb = 1;
 else if DsF = 'DFc' then DFc = 1;
 else if DsF = '   ' then do;
 DFa = 1/3;
 DFb = 1/3;
 DFc = 1/3;
end;
 if      CrF = 'CFa' then CFa = 1;
 else if CrF = 'CFb' then CFb = 1;
```

```
 else if CrF = 'CFc' then CFc = 1;
 else if CrF = 'CFd' then CFd = 1;
 else if CrF = 'CFe' then CFe = 1;
 else if CrF = 'CFf' then CFf = 1;
 else if CrF = 'CFg' then CFg = 1;
 else if CrF = 'CFh' then CFh = 1;
 else if CrF = '   ' then do;
 CFa = 1/8;
 CFb = 1/8;
 CFc = 1/8;
 CFd = 1/8;
 CFe = 1/8;
 CFf = 1/8;
 CFg = 1/8;
 CFh = 1/8;
end;
 if      FmF = 'FFa' then FFa = 1;
 else if FmF = 'FFb' then FFb = 1;
 else if FmF = 'FFc' then FFc = 1;
 else if FmF = '   ' then do;
 FFa = 1/3;
 FFb = 1/3;
 FFc = 1/3;
end;
 if      PbF = 'PFa' then PFa = 1;
 else if PbF = 'PFb' then PFb = 1;
 else if PbF = 'PFc' then PFc = 1;
 else if PbF = '   ' then do;
 PFa = 1/3;
 PFb = 1/3;
 PFc = 1/3;
end;
 if      NFA = 'NXa' then NXa = 1;
 else if NFA = 'NXb' then NXb = 1;
 else if NFA = 'NXc' then NXc = 1;
 else if NFA = 'NXd' then NXd = 1;
 else if NFA = 'NXe' then NXe = 1;
 else if NFA = 'NXf' then NXf = 1;
 else if NFA = 'NXg' then NXg = 1;
 else if NFA = 'NXh' then NXh = 1;
 else if NFA = '   ' then do;
 NXa = 1/8;
 NXb = 1/8;
 NXc = 1/8;
 NXd = 1/8;
 NXe = 1/8;
 NXf = 1/8;
 NXg = 1/8;
 NXh = 1/8;
end;
 if      PFr = 'PFa' then PFa = 1;
 else if PFr = 'PFb' then PFb = 1;
 else if PFr = 'PFc' then PFc = 1;
 else if PFr = 'PFd' then PFd = 1;
 else if PFr = 'PFe' then PFe = 1;
 else if PFr = 'PFf' then PFf = 1;
 else if PFr = 'PFg' then PFg = 1;
 else if PFr = '   ' then do;
 PFa = 1/7;
 PFb = 1/7;
```

```
 PFc = 1/7;
 PFd = 1/7;
 PFe = 1/7;
 PFf = 1/7;
 PFg = 1/7;
end;
 if      CrC = 'CCa' then CCa = 1;
 else if CrC = 'CCb' then CCb = 1;
 else if CrC = 'CCc' then CCc = 1;
 else if CrC = 'CCd' then CCd = 1;
 else if CrC = 'CCe' then CCe = 1;
 else if CrC = 'CCf' then CCf = 1;
 else if CrC = 'CCg' then CCg = 1;
 else if CrC = 'CCh' then CCh = 1;
 else if CrC = 'CCi' then CCi = 1;
 else if CrC = 'CCj' then CCj = 1;
 else if CrC = 'CCk' then CCk = 1;
 else if CrC = '   ' then do;
 CCa = 1/11;
 CCb = 1/11;
 CCc = 1/11;
 CCd = 1/11;
 CCe = 1/11;
 CCf = 1/11;
 CCg = 1/11;
 CCh = 1/11;
 CCi = 1/11;
 CCj = 1/11;
 CCk = 1/11;
end;
 if      CCR = 'CMa' then CMa = 1;
 else if CCR = 'CMb' then CMb = 1;
 else if CCR = 'CMc' then CMc = 1;
 else if CCR = 'CMd' then CMd = 1;
 else if CCR = 'CMe' then CMe = 1;
 else if CCR = 'CMf' then CMf = 1;
 else if CCR = '   ' then do;
 CMa = 1/6;
 CMb = 1/6;
 CMc = 1/6;
 CMd = 1/6;
 CMe = 1/6;
 CMf = 1/6;
end;
 if      FmC = 'FCa' then FCa = 1;
 else if FmC = 'FCb' then FCb = 1;
 else if FmC = 'FCc' then FCc = 1;
 else if FmC = '   ' then do;
 FCa = 1/3;
 FCb = 1/3;
 FCc = 1/3;
end;
 if      Cat = 'CAa' then CAa = 1;
 else if Cat = 'CAb' then CAb = 1;
 else if Cat = 'CAc' then CAc = 1;
 else if Cat = 'CAd' then CAd = 1;
 else if Cat = 'CAe' then CAe = 1;
 else if Cat = 'CAf' then CAf = 1;
 else if Cat = '   ' then do;
 CAa = 1/6;
```

```
 CAb = 1/6;
 CAc = 1/6;
 CAd = 1/6;
 CAe = 1/6;
 CAf = 1/6;
end;
 if      CFn = 'CFa' then CFa = 1;
 else if CFn = 'CFb' then CFb = 1;
 else if CFn = 'CFc' then CFc = 1;
 else if CFn = 'CFd' then CFd = 1;
 else if CFn = 'CFe' then CFe = 1;
 else if CFn = 'CFf' then CFf = 1;
 else if CFn = 'CFg' then CFg = 1;
 else if CFn = '   ' then do;
 CFa = 1/7;
 CFb = 1/7;
 CFc = 1/7;
 CFd = 1/7;
 CFe = 1/7;
 CFf = 1/7;
 CFg = 1/7;
end;
 if      Pet = 'PEa' then PEa = 1;
 else if Pet = 'PEb' then PEb = 1;
 else if Pet = 'PEc' then PEc = 1;
 else if Pet = 'PEd' then PEd = 1;
 else if Pet = 'PEe' then PEe = 1;
 else if Pet = 'PEf' then PEf = 1;
 else if Pet = '   ' then do;
 PEa = 1/6;
 PEb = 1/6;
 PEc = 1/6;
 PEd = 1/6;
 PEe = 1/6;
 PEf = 1/6;
end;
 if      McE = 'MEa' then MEa = 1;
 else if McE = 'MEb' then MEb = 1;
 else if McE = '   ' then do;
 MEa = 1/2;
 MEb = 1/2;
end;
 if      Pgl = 'PCa' then PCa = 1;
 else if Pgl = 'PCb' then PCb = 1;
 else if Pgl = '   ' then do;
 PCa = 1/2;
 PCb = 1/2;
end;
 if      MgC = 'MCa' then MCa = 1;
 else if MgC = 'MCb' then MCb = 1;
 else if MgC = 'MCc' then MCc = 1;
 else if MgC = '   ' then do;
 MCa = 1/3;
 MCb = 1/3;
 MCc = 1/3;
end;
 if      CCL = 'CAa' then CAa = 1;
 else if CCL = 'CAb' then CAb = 1;
 else if CCL = '   ' then do;
 CAa = 1/2;
```

```
 CAb = 1/2;
end;
 if      CFI = 'CIa' then CIa = 1;
 else if CFI = 'CIb' then CIb = 1;
 else if CFI = 'CIc' then CIc = 1;
 else if CFI = 'CId' then CId = 1;
 else if CFI = 'CIe' then CIe = 1;
 else if CFI = 'CIf' then CIf = 1;
 else if CFI = 'CIg' then CIg = 1;
 else if CFI = 'CIh' then CIh = 1;
 else if CFI = 'CIi' then CIi = 1;
 else if CFI = 'CIj' then CIj = 1;
 else if CFI = '   ' then do;
 CIa = 1/10;
 CIb = 1/10;
 CIc = 1/10;
 CId = 1/10;
 CIe = 1/10;
 CIf = 1/10;
 CIg = 1/10;
 CIh = 1/10;
 CIi = 1/10;
 CIj = 1/10;
end;
 if      PsF = 'PFa' then PFa = 1;
 else if PsF = 'PFb' then PFb = 1;
 else if PsF = 'PFc' then PFc = 1;
 else if PsF = 'PFd' then PFd = 1;
 else if PsF = 'PFe' then PFe = 1;
 else if PsF = 'PFf' then PFf = 1;
 else if PsF = 'PFg' then PFg = 1;
 else if PsF = '   ' then do;
 PFa = 1/7;
 PFb = 1/7;
 PFc = 1/7;
 PFd = 1/7;
 PFe = 1/7;
 PFf = 1/7;
 PFg = 1/7;
end;
 if      APF = 'APa' then APa = 1;
 else if APF = 'APb' then APb = 1;
 else if APF = '   ' then do;
 APa = 1/2;
 APb = 1/2;
end;
 if      CFM = 'CMa' then CMa = 1;
 else if CFM = 'CMb' then CMb = 1;
 else if CFM = 'CMc' then CMc = 1;
 else if CFM = 'CMd' then CMd = 1;
 else if CFM = 'CMe' then CMe = 1;
 else if CFM = 'CMf' then CMf = 1;
 else if CFM = 'CMg' then CMg = 1;
 else if CFM = 'CMh' then CMh = 1;
 else if CFM = 'CMi' then CMi = 1;
 else if CFM = 'CMj' then CMj = 1;
 else if CFM = 'CMk' then CMk = 1;
 else if CFM = 'CMl' then CMl = 1;
 else if CFM = 'CMm' then CMm = 1;
 else if CFM = 'CMn' then CMn = 1;
```

```
 else if CFM = '    ' then do;
 CMa = 1/14;
 CMb = 1/14;
 CMc = 1/14;
 CMd = 1/14;
 CMe = 1/14;
 CMf = 1/14;
 CMg = 1/14;
 CMh = 1/14;
 CMi = 1/14;
 CMj = 1/14;
 CMk = 1/14;
 CMl = 1/14;
 CMm = 1/14;
 CMn = 1/14;
end;
 if       FFt = 'FFa' then FFa = 1;
 else if FFt = 'FFb' then FFb = 1;
 else if FFt = 'FFc' then FFc = 1;
 else if FFt = 'FFd' then FFd = 1;
 else if FFt = 'FFe' then FFe = 1;
 else if FFt = 'FFf' then FFf = 1;
 else if FFt = '    ' then do;
 FFa = 1/6;
 FFb = 1/6;
 FFc = 1/6;
 FFd = 1/6;
 FFe = 1/6;
 FFf = 1/6;
end;
 if       Oft = 'OFa' then OFa = 1;
 else if Oft = 'OFb' then OFb = 1;
 else if Oft = 'OFc' then OFc = 1;
 else if Oft = 'OFd' then OFd = 1;
 else if Oft = 'OFe' then OFe = 1;
 else if Oft = '    ' then do;
 OFa = 1/5;
 OFb = 1/5;
 OFc = 1/5;
 OFd = 1/5;
 OFe = 1/5;
end;
 if       PCF = 'PBa' then PBa = 1;
 else if PCF = 'PBb' then PBb = 1;
 else if PCF = '    ' then do;
 PBa = 1/2;
 PBb = 1/2;
end;
 if       FPF = 'FPa' then FPa = 1;
 else if FPF = 'FPb' then FPb = 1;
 else if FPF = 'FPc' then FPc = 1;
 else if FPF = 'FPd' then FPd = 1;
 else if FPF = '    ' then do;
 FPa = 1/4;
 FPb = 1/4;
 FPc = 1/4;
 FPd = 1/4;
end;
 if       STv = 'STa' then STa = 1;
 else if STv = 'STb' then STb = 1;
```

```
 else if STv = 'STc' then STc = 1;
 else if STv = '   ' then do;
 STa = 1/3;
 STb = 1/3;
 STc = 1/3;
end;
 if      SFt = 'SFa' then SFa = 1;
 else if SFt = 'SFb' then SFb = 1;
 else if SFt = 'SFc' then SFc = 1;
 else if SFt = 'SFd' then SFd = 1;
 else if SFt = 'SFe' then SFe = 1;
 else if SFt = '   ' then do;
 SFa = 1/5;
 SFb = 1/5;
 SFc = 1/5;
 SFd = 1/5;
 SFe = 1/5;
end;
 if      PtF = 'PFa' then PFa = 1;
 else if PtF = 'PFb' then PFb = 1;
 else if PtF = 'PFc' then PFc = 1;
 else if PtF = '   ' then do;
 PFa = 1/3;
 PFb = 1/3;
 PFc = 1/3;
end;
 if      CPl = 'CPa' then CPa = 1;
 else if CPl = 'CPb' then CPb = 1;
 else if CPl = 'CPc' then CPc = 1;
 else if CPl = '   ' then do;
 CPa = 1/3;
 CPb = 1/3;
 CPc = 1/3;
end;
 if      MVr = 'MVa' then MVa = 1;
 else if MVr = 'MVb' then MVb = 1;
 else if MVr = '   ' then do;
 MVa = 1/2;
 MVb = 1/2;
end;
 if      Pug = 'Pga' then Pga = 1;
 else if Pug = 'Pgb' then Pgb = 1;
 else if Pug = 'Pgc' then Pgc = 1;
 else if Pug = 'Pgd' then Pgd = 1;
 else if Pug = '   ' then do;
 Pga = 1/4;
 Pgb = 1/4;
 Pgc = 1/4;
 Pgd = 1/4;
end;
 if      Arm = 'Ara' then Ara = 1;
 else if Arm = 'Arb' then Arb = 1;
 else if Arm = 'Arc' then Arc = 1;
 else if Arm = '   ' then do;
 Ara = 1/3;
 Arb = 1/3;
 Arc = 1/3;
end;
 if      CSt = 'CSa' then CSa = 1;
 else if CSt = 'CSb' then CSb = 1;
```

```
 else if CSt = 'CSc' then CSc = 1;
 else if CSt = 'CSd' then CSd = 1;
 else if CSt = '   ' then do;
 CSa = 1/4;
 CSb = 1/4;
 CSc = 1/4;
 CSd = 1/4;
end;
 if      SSt = 'SSa' then SSa = 1;
 else if SSt = 'SSb' then SSb = 1;
 else if SSt = 'SSc' then SSc = 1;
 else if SSt = '   ' then do;
 SSa = 1/3;
 SSb = 1/3;
 SSc = 1/3;
end;
 if      NSF = 'NSa' then NSa = 1;
 else if NSF = 'NSb' then NSb = 1;
 else if NSF = 'NSc' then NSc = 1;
 else if NSF = '   ' then do;
 NSa = 1/3;
 NSb = 1/3;
 NSc = 1/3;
end;
drop i;
cards;
A001 Dfb Fta Epb CHc DHc LPa NLc CFe LFb PFc CPc Sga Prk CHa Atb FHb PHa
CPe Hbb Dsb Btd DFb CFc FFb PFa NXe PFc CCa CMc FCa CAb CFb PEc MEa PCa
MCc CAa CIg PFd APa CMh FFa OFb PBa FPb STa SFb PFc CPc MVa Pgd Ara CSa
SSa NSb
A002 Dfb Fta Epb CHc DHc LPb NLc CFd LFb PFc CPb Sgb Prk CHb Atb FHb PHa
CPe Hbb Dsb Btd DFb CFc FFb PFa NXe PFc CCa CMc FCa CAb CFb PEb MEa PCa
MCc CAa CIg PFd APa CMh FFa OFb PBa FPa STa SFb PFc CPc MVa Pgc Ara CSa
SSa NSb
A003 Dfc Ftb Epb CHb DHb LPb NLc CFd LFb PFc CPb Sga Prk CHb Atb FHb PHa
CPd Hbb Dsb Btb DFb CFc FFa PFa NXe PFc CCa CMc FCb CAb CFa PEc MEa PCa
MCc CAa CIc PFd APa CMg FFa OFb PBa FPc STa SFb PFc CPc MVa Pgd Ara CSa
SSa NSb
...
;
proc print;
run;
proc corresp data=Tese out=results all dimens=189 mca;
     tables Dfl Dft EFt CHt DHt LPt NLL CFt LFt PFt CPF Seg Pro Esp CrH
AtN FHt PHt CPt Hbt DsR BtF DsF CrF FmF PbF NFA PFr CrC CCR FmC Cat CFn
Pet McE Pgl MgC CCL CFI PsF APF CFM FFt Oft PCF FPF STv SFt PtF CPl MVr
Pug Arm CSt SSt NSF;
run;
```

Apêndice C. Rotina computacional para implementação da análise de correspondência simples, no aplicativo SAS (*Statistical Analysis System Institute*, 2002), relativa a dados de caracterização de acessos do Banco de Germoplasma de *Capsicum* spp. da Embrapa Hortaliças (2007).

```
data phds;
input descritores $ 1-4 d1-d56;
label
d1 = 'Dfl' d2 = 'Dft' d3 = 'EFt' d4 = 'CHt' d5 = 'DHt' d6 = 'LPt' d7 =
'NLL' d8 = 'CFt' d9 = 'LFt' d10 = 'PFt' d11 = 'CPF' d12 = 'Seg' d13 =
'Pro' d14 = 'Esp' d15 = 'CrH' d16 = 'AtN' d17 = 'FHt' d18 = 'PHt' d19 =
'CPt' d20 = 'Hbt' d21 = 'DsR' d22 = 'BtF' d23 = 'DsF' d24 = 'CrF' d25 =
'FmF' d26 = 'PbF' d27 = 'NFA' d28 = 'PFr' d29 = 'CrC' d30 = 'CCR' d31 =
'FmC' d32 = 'Cat' d33 = 'CFn' d34 = 'Pet' d35 = 'McE'
d36 = 'Pgl' d37 = 'MgC' d38 = 'CCL' d39 = 'CFI' d40 = 'PsF' d41 = 'APF'
d42 = 'CFM' d43 = 'FFt' d44 = 'Oft' d45 = 'PCF' d46 = 'FPF' d47 = 'STv'
d48 = 'SFt' d49 = 'PtF' d50 = 'CPl' d51 = 'MVr' d52 = 'Pug'
d53 = 'Arm' d54 = 'CSt' d55 = 'SSt' d56 = 'NSF'
;
cards;
A001 Dfb Fta Epb CHc DHc LPa NLc CFe LFb PFc CPc Sga Prk CHa Atb FHb PHa
CPe Hbb Dsb Btd DFb CFc FFb PFa NXe PFc CCa CMc FCa CAb CFb PEc MEa PCa
MCc CAa CIg PFd APa CMh FFa OFb PBa FPb STa SFb PFc CPc MVa Pgd Ara CSa
SSa NSb
A002 Dfb Fta Epb CHc DHc LPb NLc CFd LFb PFc CPb Sgb Prk CHb Atb FHb PHa
CPe Hbb Dsb Btd DFb CFc FFb PFa NXe PFc CCa CMc FCa CAb CFb PEb MEa PCa
MCc CAa CIg PFd APa CMh FFa OFb PBa FPa STa SFb PFc CPc MVa Pgc Ara CSa
SSa NSb
A003 Dfc Ftb Epb CHb DHb LPb NLc CFd LFb PFc CPb Sga Prk CHb Atb FHb PHa
CPd Hbb Dsb Btb DFb CFc FFa PFa NXe PFc CCa CMc FCb CAb CFa PEc MEa PCa
MCc CAa CIc PFd APa CMg FFa OFb PBa FPc STa SFb PFc CPc MVa Pgd Ara CSa
SSa NSb
...
;
proc corresp data=phds out=results all dimens=52;
var d1-d56;
id descritores;
run;
```