

FRANCIELLE ALLINE MARTINS

**INTEGRAÇÃO DE DADOS MORFOAGRONÔMICOS, MOLECULARES E
FITOPATOLÓGICOS PARA ESTABELECIMENTO DE COLEÇÃO NUCLEAR**

Tese apresentada à
Universidade Federal de Viçosa,
como parte das exigências do
Programa de Pós-Graduação em
Genética e Melhoramento, para
obtenção do título de *Doctor
Scientiae*

VIÇOSA
MINAS GERAIS - BRASIL
2011

FRANCIELLE ALLINE MARTINS

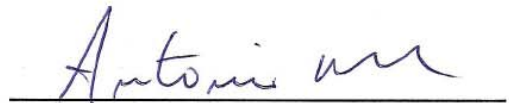
**INTEGRAÇÃO DE DADOS MORFOAGRONÔMICOS, MOLECULARES E
FITOPATOLÓGICOS PARA ESTABELECIMENTO DE COLEÇÃO NUCLEAR**

Tese apresentada à
Universidade Federal de Viçosa,
como parte das exigências do
Programa de Pós-Graduação em
Genética e Melhoramento, para
obtenção do título de *Doctor
Scientiae*

APROVADA: 17 de janeiro de 2011.



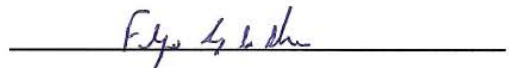
Prof. Cosme Damião Cruz
(Co-orientador)



Prof. Antônio Policarpo Souza Carneiro



Prof. Leonardo Lopes Bhering



Dr. Felipe Lopes da Silva



Prof. Pedro Crescêncio Souza Carneiro
(Orientador)

Ao meu pai
À minha mãe
Ao meu esposo
Ao meu filho
Dedico

AGRADECIMENTOS

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Genética e Melhoramento pela oportunidade de realização do curso.

Ao Conselho Nacional de Desenvolvimento científico e Tecnológico – CNPq pela concessão da bolsa de estudo.

À Universidade Estadual do Piauí pelo incentivo à conclusão do curso.

Aos professores Derly José Henriques da Silva, Cosme Damião Cruz e José Eustáquio de Souza Carneiro pela co-orientação.

Ao professor Pedro Crescêncio Souza Carneiro pela orientação, amizade e confiança.

Aos professores e amigos de curso que, de alguma forma, contribuíram para minha formação.

As amigas de todas as horas Maria Andréia e Janaina.

Aos novos amigos da Universidade Estadual do Piauí, em especial a Sorainy, Frank e Cledinaldo pela paciência e compreensão.

Aos meus amores Pedro Marcos de Almeida e Lucas Martins de Almeida pelo apoio, compreensão e amor.

Aos meus pais e irmã que sempre torceram por mim.

Por fim, agradeço a Deus pela vida e pelas oportunidades.

BIOGRAFIA

Francielle Aline Martins, filha de Gilsimar Francisco Martins e Warly Orlando Martins, nasceu em Juiz de Fora, MG, no dia 30 de abril de 1982.

Em agosto de 2004, graduou-se em Licenciatura em Ciências Biológicas e em janeiro de 2005 concluiu o Bacharelado em Ciências Biológicas, ambos pela Universidade Federal de Viçosa (UFV).

Em março de 2005 iniciou no curso de Pós-Graduação em Genética e Melhoramento de Plantas na UFV, tendo obtido o título de *Magister Science* em setembro de 2006.

Em outubro de 2006 ingressou no programa de Pós-Graduação em Genética e Melhoramento de Plantas na UFV, em nível de doutorado, submetendo-se à defesa de tese em 17 de janeiro de 2011.

SUMÁRIO

RESUMO.....	viii
ABSTRACT.....	xi
1. INTRODUÇÃO.....	1
2. REVISÃO BIBLIOGRÁFICA.....	3
2.1 Aspectos econômicos e sociais.....	3
2.2 Origem e domesticação.....	4
2.3 Recursos genéticos.....	5
2.4 Bancos de germoplasma.....	9
2.5 Diversidade Genética.....	12
2.5.1 Métodos de avaliação da diversidade genética.....	13
2.5.2 Diversidade Genética do tomateiro.....	14
2.6 Coleção nuclear.....	17
3. OBJETIVOS.....	23
4. MATERIAL E MÉTODOS.....	24
4.1 Caracterização morfoagronômica quantitativa e multicategórica.....	24
4.1.1 Caracteres quantitativos avaliados.....	26
4.1.2 Caracteres multicategóricos avaliados.....	27
4.2 Caracterização fitopatológica.....	29
4.2.1 Reação à <i>Alternaria solani</i>	29
4.2.2 Reação à <i>Pseudomonas syringae</i> pv <i>tomato</i>	30
4.2.3 Reação à begomovírus <i>Tomato yellow spot vírus</i> (ToYSV).....	30

4.3	Caracterização molecular.....	31
4.4	Diversidade Genética.....	32
4.4.1	Diversidade genética entre subamostras de tomateiro do BGH-UFV quanto à caracteres quantitativos e fitopatológicos.....	32
4.4.2	Diversidade genética entre subamostras de tomateiro do BGH-UFV quanto à caracteres multicategóricos.....	33
4.4.3	Diversidade genética entre subamostras de tomateiro do BGH-UFV quanto à caracteres moleculares.....	34
4.4.4	Comparação entre as diversidades de caracteres quantitativos/fitopatológicos, multicategóricos e moleculares.....	34
4.4.5	Integração de dados quantitativos/fitopatológicos, multicategóricos e moleculares em estudos de diversidade.....	34
4.4.5.1	Conversão dos dados quantitativos em multicategóricos.....	35
4.4.5.2	Soma algébrica de matrizes.....	37
4.4.6	Coleções Nucleares.....	37
4.4.6.1	Separação em grupos ou estratos geneticamente divergentes.....	37
4.4.6.2	Intensidade de amostragem e número de subamostras amostradas por estrato.....	38
4.4.6.3	Amostragem ou escolha das subamostras por estrato.....	38
4.4.6.4	Subcoleções avaliadas.....	39
4.4.6.5	Validação das subcoleções.....	40
5.	RESULTADOS E DISCUSSÃO.....	43
5.1	Diversidade genética entre subamostras de tomateiro do BGH-UFV quanto à caracteres quantitativos e fitopatológicos.....	43
5.1.1	Análises univariadas.....	43
5.1.2	Análises multivariadas.....	50
5.2	Diversidade genética entre subamostras de tomateiro do BGH-UFV quanto a caracteres multicategóricos e begomovírus.....	52

5.3	Diversidade genética entre subamostras de tomateiro do BGH-UFV quanto à caracteres moleculares.....	56
5.4	Comparação entre as diversidades de caracteres quantitativos, multicategóricos e moleculares.....	59
5.5	Integração de dados qualitativos, quantitativos, moleculares e fitopatológicos em estudos de diversidade.....	63
5.5.1	Conversão dos dados quantitativos em multicategóricos.....	63
5.5.2	Soma algébrica de matrizes.....	68
5.6	Coleções Nucleares.....	70
5.6.1	Separação em grupos ou estratos geneticamente divergentes.....	70
5.6.2	Número de subamostras amostradas por estrato.....	70
5.6.3	Validação das subcoleções.....	74
5.6.4	Análise gráfica da variabilidade das subcoleções.....	79
6.	CONSIDERAÇÕES FINAIS.....	91
7.	REFERÊNCIAS BIBLIOGRÁFICAS.....	93

RESUMO

MARTINS, Francielle Aline, D. Sc., Universidade Federal de Viçosa, janeiro de 2011. **Integração de dados morfoagronômicos, moleculares e fitopatológicos para estabelecimento de coleção nuclear.** Orientador: Pedro Crescêncio Souza Carneiro. Co-orientadores: Cosme Damião Cruz, Derly José Henriques da Silva e José Eustáquio de Souza Carneiro.

Objetivou-se com este estudo realizar a integração de dados de caracteres quantitativos, multicategóricos, moleculares e fitopatológicos visando à avaliação da diversidade genética de subamostras de tomateiro do Banco de Germoplasma Hortaliças da Universidade Federal de Viçosa (BGH-UFV). Foi utilizado um conjunto de dados de 67 subamostras de tomateiro do BGH-UFV caracterizadas quanto a 27 caracteres quantitativos, 34 caracteres qualitativos, 144 locos ISSR, reação a *Alternaria solani*, *Pseudomonas syringae* pv. *tomato* e ao *Tomato yellow spot vírus* (ToYSV). A avaliação da diversidade genética entre as subamostras foi realizada para cada conjunto de caracteres por meio do agrupamento de Tocher e em seguida procedeu-se a comparação entre as matrizes de dissimilaridade obtidas a partir dos dados de diferentes naturezas pelo teste Z de Mantel. Embora os valores de correlação entre as matrizes tenham sido significativos a 5% probabilidade, esses foram de baixa magnitude, não oferecendo suporte para extrapolar os resultados de um conjunto de dados para outro. Assim, duas estratégias de integração dos dados foram utilizadas: CONV – conversão dos dados quantitativos e fitopatológicos em multicategóricos, visando a obtenção de uma única matriz de dissimilaridade que contemplasse todas os caracteres

independente de suas naturezas e a SOMA - obtenção das matrizes de dissimilaridade individualmente para cada conjunto de caracteres e em seguida a soma algébrica das mesmas. Dezesete subestratégias de conversão de dados foram analisadas, dentre essas destacou-se a DEA-3 (divisão equitativa da amplitude em três classes), cujo o valor de correlação com a matriz dissimilaridade original foi de 0,782. Ao comparar as estratégias CONV e SOMA observou-se alto valor de correlação, $r=0,96$, entre as matrizes de dissimilaridade obtidas por cada uma delas. No entanto, a estratégia CONV se destacou, uma vez que permitiu maior discriminação das subamostras. Foram estabelecidas 10 subcoleções a partir das 67 subamostras de tomateiro do BGH-UFV. Essas subcoleções foram definidas pela combinação entre a natureza dos dados avaliados e intensidade de amostragem. A estratégia de amostragem utilizada foi a logarítmica estratificada, e as subamostras selecionadas por estrato foram determinadas pela análise multivariada a partir do método de agrupamento de Tocher invertido. As intensidades avaliadas foram de 20 e 30%. Para a avaliação da representatividade das subcoleções estabelecidas e a definição da que melhor representou as subamostras de tomateiro do BGH-UFV foram realizadas comparações entre a coleção inicial e as subcoleções, considerando a variância e o índice de coincidência da amplitude. A análise gráfica da variabilidade foi realizada a partir do cálculo da frequência de subamostras representadas em cada classe das características quantitativas e multicategóricas, previamente convertidas em binárias, e características moleculares. Dentre as subcoleções, a 20% de intensidade, destacou-se a subcoleção baseada na integração de dados CONV-20, por possuir maiores índices de coincidência da amplitude acompanhados de valores de variância mais adequados. A 30% de intensidade, a subcoleção MOL-30 foi tão eficiente quanto as subcoleções baseadas na integração de dados, quando considerou-se apenas o índice de coincidência da amplitude e a média das variâncias. Entretanto, a análise gráfica da variabilidade mostrou uma ligeira superioridade da subcoleção CONV-30 em manter a variabilidade, principalmente em relação aos caracteres multicategóricos. Sempre que dados de diferentes

naturezas estiverem disponíveis, deve-se priorizar o estabelecimento de coleções nucleares a partir da integração destes dados.

ABSTRACT

MARTINS, Francielle Aline, D. Sc., Universidade Federal de Viçosa, January, 2011. **Integration of morphoagronomical, molecular and phytopathological data for setting of core collection.** Adviser: Pedro Crescêncio Souza Carneiro. Co-Advisers: Cosme Damião Cruz, Derly José Henriques da Silva and José Eustáquio de Souza Carneiro.

The objective of this study was to perform the data integration of quantitative traits, multicategorical molecular phytopathological order to evaluate the genetic diversity of subsamples of tomato of Germplasm Bank of Vegetable of the Federal University of Viçosa (BGH-UFV). A data set of 67 subsamples of tomato BGH-UFV was used. It was characterized according to 27 quantitative traits, 34 qualitative characters, 144 ISSR loci, reaction to *Alternaria solani*, *Pseudomonas syringae* pv. *tomato* and Tomato yellow spot virus (ToYSV). The assessment of genetic diversity among the subsamples was performed for each set of characters by Tocher grouping and then proceeded to the comparison between the dissimilarity matrices obtained from the data of different natures by the Mantel Z test. Although the correlation values between the matrices have been significant at 5% probability, these were of low magnitude, offering no support for extrapolating results from one dataset to another. Thus, two strategies for the integration of data were used: CONV - conversion of quantitative and phytopathological data in multicategorical order to obtain a single dissimilarity matrix that encompassed all the characters regardless of their nature and SOMA - obtain the dissimilarity matrices for individual each character set and the sum of them. Seventeen sub-

strategies for converting data were analyzed; DEA-3 (fair division of amplitude in three classes) stands out among these, whose correlation value with the original dissimilarity matrix was 0.782. By comparing the strategies SOMA and CONV, it was observed high correlation coefficient, $r = 0.96$, between the dissimilarity matrices obtained by each. However, the strategy CONV stood out, since it allowed greater discrimination of the subsamples. Were established 10 subcollections from 67 subsamples of tomato from BGH-UFV. These subcollections were defined by combining the nature of data collected and sampling intensity. The sampling strategy used was the logarithm stratified, and the subsamples selected for each stratum were determined by multivariate analysis from the grouping method of Tocher reversed. The intensities were evaluated from 20 to 30%. To assess the representativeness of subcollections set and setting that best represented the subsamples of tomato BGH-UFV comparisons were made between the initial collection and subcollections, considering the variance and the index of coincidence of the amplitude. The graphical analysis of variability was performed by calculating the frequency of subsamples represented in each class of quantitative and multicategoric traits, previously converted into binary, and molecular characteristics. Among subcollections at 20% intensity, it stood out the subcollection based data integration CONV-20, for own the highest rates of coincidence of the amplitude followed by values of variance more appropriate. The 30% intensity, the subcollection MOL-30 was as efficient as the subcollections based on integration of data, when only the index of coincidence of the amplitude and the average of the variances were considered. However, the graphical analysis of variability showed a slight superiority of the subcollection CONV-30 in maintaining the variability, especially regarding the characters multicategoric. Where data from different sources are available, the establishment of core collections based on the integration of these data should be prioritized.

1. INTRODUÇÃO

A FAO (*Food and Agriculture Organization*), no ano de 1996, identificou cerca de 1300 bancos de germoplasma, preservando mais de 6,1 milhões de subamostras, dentre esses 78.000 de tomate (Fraleigh, 2006). Os bancos de germoplasma constituem uma forma de conservação *ex situ* dos recursos genéticos, sendo fundamentais para melhorar a produtividade e sustentabilidade agrícola, os quais contribuem para o desenvolvimento nacional, a segurança alimentar e a redução da pobreza (Aguilera, 2007). A potencialização do uso dos recursos genéticos em um programa de melhoramento passa, necessariamente, pela caracterização morfoagronômica, fitopatológica, entomológica e molecular da variabilidade genética retida nas coleções de subamostras registradas em um banco de germoplasma (Valois et al., 1998).

Nesse sentido, o Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa – BGH-UFV, fundado em 1966, vem trabalhando no resgate de espécies nativas ou introduzidas, preservando, documentando e disponibilizando o germoplasma, uma vez caracterizado e avaliado o seu potencial, para as condições climáticas das diversas regiões de Minas Gerais e do país (Silva et al., 2001).

A caracterização e avaliação dos recursos genéticos da coleção de *L. esculentum* Mill. do BGH-UFV vem sendo realizada de forma sistematizada. As 876 subamostras foram divididas em lotes, segundo a época de registro no BGH-UFV. Cada lote está sendo submetido à análise mediante o uso de descritores

morfológicos e agronômicos recomendados pelo International Plant Genetic Resources Institute – IPGRI (1996). Em seguida estas subamostras passam à caracterização entomológica, fitopatológica, as quais visam avaliar o comportamento das subamostras frente às principais pragas e doenças para o Brasil. Na última etapa estas subamostras são avaliadas quanto à variabilidade molecular. A partir da caracterização e avaliação é possível realizar estudos de diversidade genética, incluindo a elaboração de uma ou mais coleções nucleares (Silva et al., 2006).

Para analisar a diversidade genética, bem como avaliar os aspectos envolvidos no estabelecimento de uma coleção nuclear, geralmente os critérios utilizados são a origem geográfica (95% das coleções publicadas), características morfológicas (77%), grupos taxonômicos específicos e intraespecíficos (63%), regiões ecogeográficas (34%) e marcadores genéticos (10%) (Brown e Spillane, 1999).

Até o momento, tem-se observado a realização de estudos de diversidade genética e o estabelecimento de coleções nucleares com base em grupos isolados de caracteres, sejam quantitativos, fitopatológicos, multicategóricos ou moleculares. A reunião de todos os dados em um único estudo tem sido dificultada, em parte por existirem poucos bancos de germoplasma que possuem uma caracterização do seu germoplasma tão detalhada. Outro fator que interfere nesse tipo de estudo é a ausência de uma metodologia de análise eficiente e disponível para a comunidade científica.

Assim, objetivou-se a integração de caracteres quantitativos, fitopatológicos, multicategóricos e moleculares em estudos de diversidade genética do tomateiro bem como o estabelecimento de subcoleções nucleares representativas.

2. REVISÃO BIBLIOGRÁFICA

2.1 Aspectos econômicos e sociais

O tomate é uma hortaliça muito popular e cultivada em todo mundo, que participa da culinária típica de diversas etnias. A ampla aceitação mundial se deve à variedade de tipos de frutos, com diversos formatos, tamanhos, cores e sabores (Consuegra et al., 2000).

No mundo são produzidos em torno de 126 milhões de toneladas por ano de tomate (FAO, 2007), sendo os maiores produtores China e Estados Unidos. Em 2007, a produção brasileira de tomate foi de 3,2 milhões de toneladas, em área cultivada de 54,912ha e produtividade média de 53,8 t/ha (Agrianual, 2008). Maior parte da produção brasileira (70%) é destinada ao consumo *in natura* (EMBRAPA, 2004), enquanto apenas 21% da produção americana é destinada a este fim, sendo o restante processado pelas indústrias de alimento (Fontes e Silva, 2002) na preparação de conservas, molhos e ketchups. O aumento da demanda do tomate industrial tem impulsionado a expansão da produção no mundo, enquanto a taxa de crescimento do tomate para consumo *in natura* acompanha o crescimento da população.

No Brasil, o tomateiro ocupa o segundo lugar em volume de produção de hortaliças, ficando atrás apenas da batata (Silva et al., 2003). Além de sua importância econômica, a cadeia produtiva do tomate desempenha papel social

relevante, pois possui mais de 10 mil produtores e envolve 60 mil famílias de trabalhadores, cujo efetivo é de mais de 200 mil pessoas (Tavares, 2003).

A produção do fruto está concentrada no centro-oeste e sudeste do país, sendo o estado de Goiás o maior produtor seguido de São Paulo e Minas Gerais, responsáveis por mais de 60% da produção. Com menor expressão estão o Rio de Janeiro e Bahia, variando de pequenas áreas em cultivo protegido até grandes extensões, destinadas à industrialização. Em alguns casos, principalmente para a indústria, a colheita é realizada mecanicamente; no outro extremo, para o tomate de mesa, a colheita é escalonada e manual (Rick, 1995).

O bom estágio econômico brasileiro associado ao nível tecnológico da produção às reservas de água e solo favorável tem fornecido não só condições para a expansão da produção do tomate de mesa, mas principalmente para a cultura se firmar como abastecedora de derivados de tomate processado dos mercados latino-americano e da Ásia (Camargo Filho e Alves, 2003).

2.2 Origem e domesticação

O tomate (*Lycopersicon esculentum*) tem o seu principal centro de diversidade em um estreito cinturão ao longo da região andina do Equador e do Peru. Diferentemente da maioria das espécies cultivadas, o tomate não foi domesticado no seu local de origem. A hipótese comumente aceita para a domesticação do tomate cultivado é que a *L. esculentum subsp. cerasiforme* espalhou como uma erva daninha a partir da região andina para o México, onde foi domesticado (Rick, 1995). Sua disseminação foi facilitada pelo reduzido tamanho de suas sementes que podiam ser facilmente carregadas (Hays e Hays, 1973).

Com a colonização, o tomate foi levado pelos espanhóis para ser domesticado na Europa no século XVI, e em seguida levado para a Itália e Inglaterra. Inicialmente o tomate foi considerado como planta ornamental, e não era consumida na alimentação, pelo temor a toxicidade (Camargo et al., 2006). Por volta de 1711 o tomate foi levado para os Estados Unidos, onde também foi

cultivado como planta ornamental e não era consumido por ser considerado venenoso. Até que em 1820, Robert Gibbon Johnson, teve a ousadia de comer tomates em público, desafiando a crença de que quem os consumisse morreria intoxicado. Só depois desse episódio o tomate começou a ser cultivado e consumido, mas ainda em pequena escala (Smith, 1994).

Depois da guerra civil (1861-1865), a cultura do tomateiro para alimentação foi divulgada para muitas áreas do mundo, inclusive no Brasil, um maior número de pessoas nos Estados Unidos aceitou os tomates como parte de sua alimentação e, por volta de 1892, já havia uma grande demanda para seu consumo (Hays e Hays, 1973).

Durante a sua evolução e domesticação o *L. esculentum* foi submetido a vários 'gargalos' genético impostos pela autopolinização, efeito fundador, seleção natural e artificial, e a extremos de endogamia com um número limitado de genótipos para cruzamentos, especialmente na Europa e na América do Norte (Rick 1991) de forma que a cultura hoje apresenta uma estreita base genética. Esta situação pode ser revertida a partir da utilização de novos genitores, de modo a contribuir para a ampliação da base genética do tomateiro cultivado (Stevens e Rick, 1986). Para tanto é necessário que o melhorista conheça os recursos genéticos disponíveis, ou seja, o germoplasma da espécie com a qual trabalha, pois, assim a introdução de novos genitores nos programas de melhoramento poderá ser realizada de forma direcionada e a vulnerabilidade resultante do estreitamento da base genética ser evitada com o aumento da variabilidade (Silva et al., 2001).

2.3 Recursos genéticos

O melhoramento genético das plantas cultivadas requer a disponibilidade da variabilidade genética disponível na natureza e em áreas cultivadas, resultantes de hibridação natural ou mutação e mantidas pela seleção natural. Esta variabilidade natural permaneceu sem o controle do homem até o final do século dezenove, quando foram iniciados os métodos científicos de melhoramento

genético das plantas. Entretanto, os sistemas de cultivos tiveram um papel importante na formação de raças indígenas locais de plantas cultivadas, assim chamadas “formas primitivas”, resultantes de uma intensa seleção massal. (Goedert, 2006).

Em anos recentes, tem-se evidenciado que grande parte da variabilidade genética das raças locais está em perigo de extinção e que, em muitas regiões, já está desaparecida, em consequência da pressão das cultivares melhoradas e de práticas agrícolas avançadas. Este problema agrava-se, quando é somado ao fato de que estas raças locais estão desaparecendo, sem terem sido caracterizadas e avaliadas para eventual utilização em programas de melhoramento (Goedert, 2006).

À medida que as raças locais são substituídas, os alelos responsáveis pela adaptação e resistência a doenças e pragas estão sendo perdidos, expondo a comunidade científica à seguinte realidade: por um lado, a necessidade de conservação da variabilidade genética das raças locais como valiosas fontes de variabilidade genética e, por outro, a necessidade premente da criação de novas variedades das espécies econômicas com alta produtividade para atender a crescente necessidade de alimentos. Tendo presente esta realidade e a conscientização de que a conservação dos recursos genéticos autóctones e exóticos desempenha função vital para a sobrevivência da humanidade, procura-se compatibilizar a conquista de elevada produção biológica com a manutenção dos recursos genéticos ainda existentes (Goedert, 2006).

A dependência da humanidade por recursos genéticos para a continuação da vida no planeta é fato indiscutível, por isso, a sua conservação é estratégica para garantir por muitas décadas, a sobrevivência das sementes de espécies de interesse socioeconômico atual e potencial, assegurando a manutenção das fontes básicas para a alimentação e para agricultura (Goedert, 2006).

Nesse sentido, foi criada a Organização das Nações Unidas para a Alimentação e Agricultura – FAO com o objetivo de realizar esforços, em nível internacional, de combate à fome. Com a participação de mais de 180 países, entre os quais o Brasil, a FAO atua como um fórum global neutro visando

harmonizar as políticas ligadas à agricultura e garantir o acesso aos recursos genéticos (Wetzel e Macedo, 2006).

Uma das primeiras estratégias internacionais consistentes relativas à importância do manejo e conservação dos recursos genéticos foi a criação da Comissão Internacional de Recursos Genéticos Vegetais (*International Board for Plant Genetic Resources* – IBPGR) durante a Conferência Técnica Internacional em biologia realizada pela FAO, em 1974 (Goedert, 2007), visando promover, supervisionar e coordenar uma rede internacional de centros de recursos genéticos que atendessem as suas atividades inerentes (Wetzel e Macedo, 2006).

No início de 1990, tornou-se evidente a necessidade de realização de uma conferência intergovernamental para conhecer o progresso, identificar problemas e oportunidades, e ajudar no direcionamento para futuras atividades de conservação e utilização dos recursos genéticos de plantas para alimentação e agricultura. Assim, em 1992, a importância dos recursos genéticos vegetais foi reconhecida na Conferência das Nações Unidas do Meio Ambiente e Desenvolvimento, quando 154 países aprovaram a Convenção sobre Diversidade Biológica – CDB. Desde essa data, mais de 175 países já ratificaram o acordo (Goedert, 2007). Os recursos genéticos, até então considerados patrimônios da humanidade, com o advento da CDB passaram a estar sujeitos a soberania do Estado. Este tratado ambiental internacional, além de alterar a natureza jurídica dos recursos genéticos, acoplou a conservação da biodiversidade à utilização sustentável e à partilha de benefícios gerados pelo uso e exploração dos recursos (Granja et al., 1999).

Logo após a entrada em vigor da CDB, o IBPGR foi transformado no *International Plant Genetic Resources Institute* – IPGRI do *Consultative Group on International Agricultural Research* – CGIAR, em 1994, com mandato de fomentar a conservação e o uso da diversidade genética, atuando em todos continentes (Wetzel e Macedo, 2006). Desde então, progresso significativo tem sido alcançado na coleta e conservação de espécies cultivadas e, conseqüentemente, grandes coleções de germoplasma foram estabelecidas (Abadie et al., 2000). O IPGRI das Américas, como é conhecido, está hospedado pelo *International Center for*

Tropical Agriculture – CIAT, na Colômbia e tem como objetivo auxiliar nas diversas atividades de recursos genéticos nas Américas (Wetzel e Macedo, 2006). A modelo do CIAT, outros centros ou institutos foram organizados, como o *International Maize and Wheat Improvement Centre* – CIMMYT, no México; o *International Crops Research Institute for the Semi-Arid tropics* - ICRISAT, na Índia; o *Internacional Rice Research Institute* – IRRI, nas Filipinas e o *International Institute of Tropical Agriculture* - IITA, na Nigéria.

No âmbito do compromisso global, em 2001, foi assinado um Tratado Internacional sobre Recursos Genéticos Vegetais para a Alimentação e Agricultura. Este Tratado visa facilitar o intercâmbio de germoplasma dos produtos agrícolas que são básicos para a alimentação humana. Mais recentemente, em 2006, foi criado o Fundo Mundial para a Diversidade de Espécies (*The Global Crop Diversity Trust*) com a participação da FAO, IPGRI e CGIAR. Este Fundo tem como objetivo assegurar a conservação e a disponibilidade dos recursos genéticos visando manter a segurança alimentar e uma agricultura sustentável (Wetzel e Macedo, 2006).

A manutenção dos recursos genéticos e de suas fontes é fundamental para a sobrevivência da humanidade, seja nos aspectos sociais, como a segurança alimentar que passa desde os incrementos de produção para garantir comida ao mundo com população crescente, ao desenvolvimento de cultivares resistentes ou tolerantes aos azares possíveis das mudanças climáticas. Ou, seja nos aspectos utilitários em que existe a necessidade de expansão da matriz bioenergética para a produção de biocombustíveis, cuja tecnologia de produção já vem sendo dominada pelo país.

2.4 Bancos de germoplasma

Dentre as primeiras coleções de germoplasma formadas, a mais significativa foi iniciada no século XX, entre os anos de 1920 e 1930 pelo All-Union Research Institute of Applied Botany and New Crops de St. Petersburg sob a coordenação de Nikolai Ivanovik Vavilov que atualmente chama-se Nikolai I. Vavilov All-Russian Scientific Research Institute of Plant Industry (VIR). Hoje, o instituto é considerado o primeiro banco de sementes e sua coleção uma das maiores de germoplasma vegetal, mais de 380.000 genes foram identificados representando a variabilidade de 2.500 espécies de plantas (Mattedi, 2009).

No Brasil, a preocupação com a coleta e preservação de recursos genéticos teve impulso decisivo na década de 70 com a criação da Embrapa Recursos Genéticos e Biotecnologia-CENARGEN. Desde então, a Embrapa, como instituição do país responsável pela pesquisa e desenvolvimento do agronegócio, vêm se esforçando para implementar as diretrizes globais de recursos genéticos, enfatizando o enriquecimento, a caracterização, a conservação e uso do germoplasma em suas diversas formas, respaldadas por um sistema organizado de documentação e informação dos recursos genéticos trabalhados. O CENARGEN, nos seus trinta e seis anos de existência, construiu estruturas físicas adequadas e implantou um sistema eficiente para a conservação a médio e longo prazo, das coleções de base de germoplasma semente e *in vitro*. Este sistema destaca-se no mundo inteiro, não só pela forma organizada em que faz a conservação *ex situ*, mas principalmente, por estar associado a uma rede de Bancos Ativos de Germoplasma, localizados em sua maioria, nas Unidades de pesquisa da Empresa, distribuídas no território nacional (Goedert, 2006).

No CENARGEN, a conservação em longo prazo de germoplasma semente é desenvolvida por meio de um Banco Base de Germoplasma Semente (Colbase), criado em 1976, o qual é constituído por câmaras frias onde estão armazenadas cerca de 97 mil subamostras de germoplasma semente nativos e exóticos, de 223 gêneros e 901 espécies. Fazem parte da Coleção de Base cerca de 2 mil subamostras, mantidas *in vitro*, de material de propagação vegetativa de várias

espécies importantes para a agricultura brasileira, além de espécies silvestres de uso potencial (Goedert, 2006). No ano de 2008, o banco de germoplasma do CENARGEN ocupou o quarto lugar no ranking entre os maiores bancos de germoplasma do mundo, com 150 mil subamostras (EMBRAPA, 2008).

Em todo o Brasil estima-se a existência de 350 bancos de germoplasma, localizados em sua maioria, nas Unidades de pesquisa da Embrapa e Universidades Federais, distribuídas no território nacional. Dentre as coleções destaca-se em maior número de subamostras a coleção internacional de cevada, as coleções de feijão, arroz, soja e trigo.

Na Universidade Federal de Viçosa (UFV) a preocupação com a preservação de germoplasma foi anterior a criação do CENARGEN, pois junto a Fundação Rockefeller, a UFV iniciou sua atividade de coleta e preservação em 1964, o que culminou com a criação do Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH-UFV) em 1966. O BGH-UFV foi criado com a finalidade de resgatar espécimes de espécies nativas ou introduzidas, preservar, documentar e disponibilizar germoplasma de outras regiões do globo, uma vez caracterizado e avaliado seu potencial para as condições climáticas das diversas regiões de Minas Gerais e do Brasil (Silva et al., 2006).

Os recursos genéticos do BGH - UFV representam mais de 50 anos de coleta, pois esta atividade, intermitente, iniciou-se em 1964 e mantém-se ativa até os dias atuais. O BGH – UFV também realizou intercâmbio de germoplasma com outros países, principalmente na década de 60, quando foram introduzidos 737 subamostras de hortaliças de outros países (Silva et al., 2001). Atualmente, o BGH - UFV possui 7480 subamostras, com 25 famílias e 106 espécies. As famílias com maiores representações são: Solanaceae, 44,21%; Leguminosae, 16,83%; Cucurbitaceae, 15,70%, as demais somam 23,26% (Silva et al., 2006).

O Banco de Germoplasma mantém diversas atividades de rotina e de pesquisa. Avaliação da germinação, multiplicação e regeneração das subamostras fazem parte das atividades de rotina do banco. Entre as atividades de pesquisa o BGH – UFV mantém processos contínuos de caracterização e avaliação dos recursos genéticos que tem sob seu controle (Silva et al., 2006).

Dentre as espécies encontradas no BGH-UFV, destaca-se *Lycopersicon esculentum* Mill. com cerca de 876 subamostras registradas, diversas destas foram coletados a partir de agricultores de diferentes partes do país. Assim estas subamostras passaram por diferentes intensidades de seleção humana e natural visando à adaptação e por isto podem ter diversos alelos do tipo comum e localmente distribuído, e raro e localmente distribuído. Por esta razão é possível encontrar novos alelos para serem utilizados em programas de melhoramento relacionados à produção, qualidade, resistência e tolerância à estresses bióticos e abióticos (Silva et al., 2006).

Além do aspecto prático de seleção de novos genes, a caracterização visa identificar subamostras duplicadas dentro de uma coleção, em função de sinônimas e diferentes doadores do mesmo germoplasma. A presença de tais duplicatas onera as atividades por consumirem dinheiro e recursos humanos para a conservação do mesmo germoplasma. Assim a caracterização permite não só a utilização dos recursos genéticos armazenados em bancos de germoplasma como também a redução dos custos da manutenção do banco pela redução do número de subamostras a serem conservadas (Valois et al., 1998).

O aumento do uso de germoplasma em programas de melhoramento e por comunidades rurais está relacionado à disponibilidade de informações sobre as subamostras mantidas nas coleções ou bancos, geradas a partir de estudos morfoagronômicos e de divergências genotípica e fenotípica, permitindo o maior conhecimento das subamostras preservadas e a possibilidade de uma ampla utilização desses (Schuelter, 1996; Bento et al., 2007).

Nesse sentido, desde 2001, a caracterização e avaliação dos recursos genéticos da coleção de *L. esculentum* Mill. do BGH-UFV vem sendo realizada de forma sistematizada, sobretudo por estudantes de iniciação científica e no desenvolvimento de dissertações e teses de pós-graduação, sob orientação do professor Derly José Henriques da Silva. As 876 subamostras foram divididas em lotes segundo a época de registro no BGH-UFV. Cada lote está sendo submetido à análise mediante o uso de descritores morfológicos e agrônômicos recomendados pelo IPGRI (1996). Em seguida estas subamostras passam à

caracterização entomológica e fitopatológica, que visam avaliar o comportamento dessas frentes às principais pragas e doenças para o Brasil. Na última etapa estas são avaliadas quanto à variabilidade molecular e ao final de todo este processo tem-se a possibilidade de construir uma coleção nuclear com base nos diferentes tipos de variabilidade existente em uma coleção.

Deste trabalho, mais de 400 subamostras de tomateiro do BGH-UFV já foram caracterizadas e avaliadas nos últimos anos pelo Núcleo de Estudos em Olericultura da UFV, e parte dessa caracterização e avaliação encontra-se disponível na rede internacional de computadores no site www.ufv.br/bgh.

2.5 Diversidade genética

A caracterização e análise da diversidade genética do germoplasma de uma espécie têm importância fundamental nos programas de melhoramento, pois a partir das avaliações é possível identificar genes de interesse agrônomo e inseri-los num programa de cruzamento com as cultivares já desenvolvidas de modo a transferir o gene de interesse (Mohammadi e Prasanna, 2003).

Outro aspecto relevante é a identificação de diversas combinações de parentais para obtenção de progênies segregantes com o máximo de variabilidade genética para posterior seleção (Barrett and Kidwell, 1998) ou até mesmo estudos de mapeamento de QTLs.

Análise da diversidade genética das coleções de germoplasma pode facilitar a classificação das subamostras, e identificação de subconjuntos com potencial de utilização para fins específicos de cruzamento. Ênfase significativa tem sido dada a uma análise abrangente da diversidade genética em diversas culturas, incluindo as grandes culturas como o trigo (*Triticum aestivum* L.), arroz (*Oryza sativa* L.), milho (*Zea mays* L.), cevada (*Hordeum vulgare* L.), e soja [*Glycine max* (L.) Merr.] (Mohammadi e Prasanna, 2003).

O estudo da diversidade genética é o processo pelo qual variações individuais ou entre grupo de indivíduos ou populações são analisadas por um específico método ou combinações de métodos. As análises envolvem medidas

numéricas e em muitos casos, combinações de diferentes tipos de variáveis (Mohammadi e Prasanna, 2003). Vários grupos de dados têm sido utilizados pelos pesquisadores para a análise de diversidade de plantas cultivadas; os mais importantes entre esses são os dados de pedigree (Bernardo, 1993; Messmer et al., 1993; van Hintum e Haalman, 1994) ou passaporte (Bar-Hen et al., 1995), que frequentemente são incompletos ou limitados (Berloo et al., 2008); dados morfológicos (Marim et al., 2002), que embora seja prática e intuitiva sua obtenção são submetidos a influência ambiental e pressão de seleção durante domesticação e melhoramento o que dificulta as interpretações dos resultados do estudo de diversidade baseada em tais caracteres (Berloo et al., 2008); dados bioquímicos, obtidos pela análise de isoenzimas (Hamrick and Godt, 1997), e dados de marcadores de DNA, que permitem maior diferenciação de genótipos (Franco et al. 2001; Zhao et al. 2005).

O conhecimento do grau de variabilidade genética, por meio dos estudos de divergência, torna-se vantajoso no processo de identificação de novas fontes de genes de interesse (Amaral Júnior e Thiébaud, 1999). Outra vantagem é o fato de que, por meio da diversidade genética, pode-se indicar progenitores geneticamente distantes para cruzamentos onde se procure obter o efeito heterótico na geração híbrida e maior probabilidade de recuperação de segregantes superiores em gerações avançadas (Amaral Júnior e Thiébaud, 1999; Cruz et al., 2004). Em outras situações, estudos sobre diversidade genética têm sido realizados com o intuito de identificar grupos de cultivares com maior similaridade, visando à formação de multilinhas (Cruz et al., 2004).

2.5.1 Métodos de avaliação da diversidade genética

A divergência genética tem sido avaliada por meio de técnicas biométricas, baseadas na quantificação da heterose, ou por processos preditivos. Dentre os métodos fundamentados em modelos biométricos, que se destinam à avaliação da divergência dos progenitores, citam-se as análises dialélicas, que avaliam tanto a

capacidade específica quanto a heterose manifestada nos híbridos (Cruz et al., 2004).

Na predição da divergência genética, vários métodos multivariados podem ser aplicados. Dentre eles, citam-se a análise por componentes principais, variáveis canônicas e os métodos aglomerativos (Cruz et al., 2004).

Os métodos preditivos de estudo da diversidade genética baseiam-se em diferenças agronômicas, morfológicas, fisiológicas ou moleculares, quantificando-as em alguma medida de dissimilaridade que expressa o grau de diversidade genética entre os genótipos. As características avaliadas em estudos de diversidade genética podem ser classificadas, de acordo com a variação apresentada em contínua/quantitativa ou discreta/multicategórica ou binária.

De maneira geral, tem se observado a realização de estudos de diversidade genética com base em caracteres quantitativos, ou qualitativos, ou moleculares, isoladamente. A reunião de todos os dados em um único estudo tem sido dificultada, em parte por existirem poucos bancos de germoplasma que possuem uma caracterização do seu germoplasma tão detalhada, em razão da limitação dos recursos financeiros e humanos. Outro fator que interfere nesse tipo de estudo é a ausência de uma metodologia de análise eficiente e disponível para a comunidade científica (Franco et al., 1997b).

2.5.2 Diversidade genética do tomateiro

A variabilidade do tomateiro vem sendo estudada, sobretudo na busca por fontes de resistência a pragas e doenças. Barona et al. (1989) caracterizaram parte do banco de germoplasma da Colômbia com o objetivo de selecionar subamostras de *Lycopersicon hirsutum* como fontes de resistência a *Tuta absoluta* para serem utilizados em programas de melhoramento visando incorporar o gene de resistência a essa praga em cultivares comerciais.

Egashira et al. (2000), estudando a variabilidade das espécies *L. peruvianum* e *L. chilense*, espécies relacionadas ao tomateiro cultivado, concluíram que estas espécies têm as duas maiores variabilidades dentro do

gênero *Lycopersicon* e que a base genética dessas espécies é claramente diferente da base da espécie mais relacionada *L. esculentum*. Desta forma, a exploração destas espécies em programas de melhoramento poderá ampliar a variabilidade para a espécie cultivada.

Buscando fontes de variabilidade para aroma e sabor dentro da espécie cultivada, Mata et al. (2000) avaliaram 26 subamostras de tomateiro do banco de germoplasma da Universidade Politécnica de Valência – Espanha. Dentre as subamostras avaliadas observaram a existência de variabilidade para as características: sólidos solúveis, acidez total e compostos relacionados ao aroma, de forma que algumas dessas subamostras puderam ser utilizadas como fonte de genes para aumentar os valores desses caracteres nas cultivares.

Com o intuito de aumentar a quantidade de carotenóide e vitamina C em tomates comerciais, Adalid et al. (2008) estudaram a diversidade entre 11 genótipos de *Lycopersicon esculentum* var. *cerasiforme*, seis variedades comerciais de *L. esculentum*, três mutantes com elevada acumulação de pigmentos e um híbrido comercial, sendo que o genótipo UPV 17790 destacou-se por possuir quatro vezes mais o teor de licopeno do híbrido e 1,35 do mutante.

Diferença significativa foi observada entre um grupo de “landraces” e cultivares comerciais de tomate no sul do Brasil quando analisadas com marcadores RAPD, indicando que as “landraces” podem ser uma fonte de variação na base genética de programas de cruzamento (Carelli et al., 2006)

Karasawa et al. (2005) caracterizaram 70 subamostras que compõem o Banco de Germoplasma de *L. esculentum* da Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), utilizando 27 descritores do IPGRI com objetivo de quantificar a divergência genética entre os mesmos, por meio de métodos de agrupamento. Os dados obtidos revelaram elevada concordância na formação de grupos entre as subamostras estudadas, tanto pelo método de Tocher quanto pelo Vizinho Mais Próximo. Concordância entre métodos de agrupamento foi verificada para pimentas e pimentões (Sudré et al., 2005), e feijão de vagem (Abreu et al., 2004). O método do Vizinho Mais Próximo possibilitou a formação de grupos, de acordo com o número de dias para germinação. Considerando que a análise

multivariada possibilita a predição da heterose, alguns cruzamentos foram sugeridos pelos autores, seguindo-se o princípio de se cruzar as subamostras mais distantes e com melhores características agrônômicas (Sudré et al., 2005).

O advento dos marcadores moleculares estimulou o estudo da diversidade em plantas, com tomate alguns trabalhos têm sido realizados desde a década de 90 usando marcadores como RFLP (Miller e Tanksley 1990), RAPD (Williams e St. Clair 1993; Villand et al. 1998; Noli et al. 1999; Carelli et al. 2006), AFLP (Park et al. 2004; Berloo et al., 2008), SSR (Alvarez et al., 2001; Bredemeijer et al. 2002) e ISSR (Aguilera, 2007).

Os estudos de diversidade conduzidos a partir do BGH-UFV já permitiram a observação da existência de subamostras com BRIX superior ao das testemunhas comerciais “Fanny”, “Débora Plus” e “Santa Clara”, característica de interesse do melhorista, sobretudo ao considerar a aceitação do fruto para o mercado consumidor. O estudo da diversidade permitiu, ainda, a observação de possíveis duplicatas, pois um grupo de 9 subamostras foram similares entre si, em todas as avaliações (Silva et al., 2006).

Estudos da dissimilaridade de 34 subamostras de tomateiro, utilizando as características da fase vegetativa e de produção permitiram, pelo método de Tocher, verificar a consistência dos grupos formados, gerando 12 classes diferentes. Os dois primeiros componentes principais foram suficientes para explicar cerca de 86% da variação total disponível entre os subamostras. As características menos importantes foram número de folhas definitivas e diâmetro do hipocótilo (Marim et al., 2002).

As subamostras de *Lycopersicon* conservadas no BGH-UFV têm sido utilizados também na busca de fontes de resistência a pragas e doenças (Silva et al., 2001), tendo sido identificado até o momento quatro boas fontes de resistência à *Fusarium oxysporum f. lycopersici* (raça 1) e à mancha de *Stemphylium solani*, simultaneamente (Matsuoka e Chaves, 1973), quatro fontes de resistência a *Tuta absoluta* (Suinaga et al., 2003), uma para *Peper yellow mosaic vírus* (PepYMV) (Juhász et al., 2006), uma para resistência ao begomovírus *Tomato yellow spot virus* (ToYSV) (Aguilera, 2007), duas para *Phytophthora infestans* (Ribeiro et al.,

2006), duas para *Alternaria solani* (Rodrigues et al., 2006a), duas para *Ralstonia solanacearum* (Rodrigues et al., 2006b), seis para *Pseudomonas syringae* (Soares et al., 2007).

O estudo da variabilidade genética de subamostras de tomateiro, com base na avaliação da eficiência fotossintética, partição de fotoassimilados e produção, realizado por Flores (2007) demonstrou a possibilidade de se ter plantas produtivas com menores índices de área foliar que as cultivares atuais, demandando assim menor necessidade de insumos por unidade de planta.

2.6 Coleção Nuclear

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 modernos e a exploração agrícola tecnificada (Oliveira, 2007).

Na década de 1980, o estímulo financeiro dado pelo IBPGR para as atividades de recursos genéticos, resultou no crescimento das coleções, que não eram totalmente, ou mesmo parcialmente caracterizadas ou avaliadas. Esse crescimento no número de subamostras e a falta de caracterização e avaliação dos materiais prejudicaram a qualidade das coleções, que estavam sendo usadas muito aquém do que deveriam (Holden, 1984; Marshall, 1989).

A ênfase dada na importância de preservar o germoplasma de culturas importantes tem levado à formação e manutenção de coleções grandes (Frankel e Bennett, 1970), a acessibilidade e utilidade de uma coleção é inversamente relacionada com o seu tamanho (Frankel e Soulé, 1981). Com o objetivo de melhorar a utilização e acessibilidade e minimizar as dificuldades de manutenção Frankel (1984) propôs pela primeira vez o conceito de “coleção nuclear”.

Uma coleção nuclear (CN) consiste num conjunto de subamostras derivadas de uma coleção de germoplasma escolhido para representar a máxima variabilidade genética da coleção inicial ou base (CI) com o mínimo de

redundância (Brown e Spillane, 1999; Vilela-Morales et al., 1997; Cordeiro e Abadie, 2007).

O estabelecimento de uma coleção nuclear é uma estratégia que prioriza e concentra atividades de caracterização e avaliação, formando uma base de informação mais completa das subamostras, melhorando de forma efetiva a conservação, acessibilidade e uso de recursos genéticos (Cordeiro e Abadie, 2007).

A montagem de uma CN, independente da espécie, é basicamente um exercício de amostragem que tenta assegurar a presença do maior número de alelos da CI para utilização pelos melhoristas da espécie. Entretanto, isto não necessariamente implica que todos os alelos têm que ser incluídos na CN. Isto ocorre, principalmente, por duas razões: a impossibilidade prática de incluir todos os alelos em uma amostra reduzida (Brown, 1989) e o fato de que todos os alelos não têm igual valor de uso no futuro (Allard, 1992).

É improvável que os alelos que estão em baixa frequência na maioria das subamostras, embora ocasionalmente afetem a qualidade de outras características de importância agrônômica, tenham alguma importância no futuro. Consequentemente, as estratégias de amostragem devem assegurar uma razoável inclusão dos alelos mais frequentes os quais estão relacionados com a adaptação geral (alta frequência e amplamente distribuídos) e adaptação específica (frequência alta ou intermediária, mas com localização restrita). De forma, que a estratégia de amostragem para obter a CN, deve maximizar a diversidade na amostra e reduzir a redundância de genótipos idênticos, o que significa que as frequências das subamostras nas classes mais representativas devem ser reduzidas, enquanto que as frequências dos tipos mais raros devem ser aumentadas. Em ambos os casos, a coleção de amostra deve incluir a maior diversidade genotípica possível (Oliveira, 2007).

A amostragem de uma coleção de germoplasma para estabelecer uma CN requer um esforço integrado envolvendo curadores, melhoristas e geneticistas para definir o tamanho da CN e escolher os critérios de classificação para as subamostras. A definição de estratégias de alocação do esforço amostral para

procurar maximizar a diversidade genética capturada na CN é também importante (Abadie et al., 2005). Nesse contexto, Johnson e Hodgkin (1999) sugeriram que cerca de 10% das subamostras devem conter 70% da diversidade genética da coleção inteira; na prática as proporções variam de 5 a 20% das subamostras e de 70 a 90% da diversidade (Araújo, 2008).

Para analisar a diversidade genética, bem como avaliar os aspectos envolvidos na implementação de uma coleção nuclear, geralmente mais de um critério pode ser utilizado. Os mais usados são: origem geográfica (95% das CN publicadas), características morfológicas (77%), grupos taxonômicos específicos e intra-específicos (63%), regiões ecogeográficas (34%) e marcadores genéticos (10%) (Brown e Spillane, 1999).

Outra etapa importante no estabelecimento de uma coleção nuclear é a determinação da proporção com que cada estrato será representado. Brown (1989) propôs três procedimentos baseados no tamanho do grupo: estratégia constante (C), proporcional (P) e logarítmica (L). A estratégia Constante (C) ignora o tamanho e o número de subamostras por estrato, propondo que um número igual de subamostras seja tomado em cada estrato. A estratégia proporcional (P) aloca a entrada por grupo em proporção ao número de subamostras em cada estrato. A estratégia logarítmica (L) propõe alocar o número de subamostras proporcional ao logaritmo do número de subamostras em cada estrato. As três estratégias fornecem diferentes números de entrada em diferentes estratos, e geralmente as estratégias proporcional e logarítmica têm sido amplamente usadas, em contraste com a estratégia constante (Van Hintum et al., 2000).

O passo final para estabelecer uma coleção nuclear é a escolha das subamostras. As subamostras selecionadas deveriam ser aquelas que melhor representam o grupo com importância para a coleção. Isso não é somente com relação à diversidade genética no grupo, mas também com respeito a outras considerações, tais como qualidade da documentação das subamostras, disponibilidade de sementes ou papel importante em programas de melhoramento. Marita et al. (2000) desenvolveram um algoritmo para estabelecer uma coleção nuclear que representa a diversidade genética máxima das subamostras baseado

em estimativas de distância genética. Uma estratégia alternativa para a seleção das subamostras componente da CN por meio da técnica de agrupamento de Tocher com critério de aglomeração inverso foi proposta por Vasconcelos et al. (2007). Esta consiste na formação de um único grupo, a partir do agrupamento de subamostras com maior dissimilaridade, que corresponde à CN. Esta estratégia tem sido empregada com sucesso no estabelecimento de coleções nucleares (Vasconcelos et al., 2007; Oliveira, 2007)

O critério aleatório para seleção das subamostras que compõem a CN tem sido usado somente quando não se dispõe de informações para determinar a escolha das subamostras para representar o estrato (Brown, 1989; Cordeiro e Abadie, 2007).

Existiam cerca de 50 coleções nucleares de vários tipos de cultivos (cereais, grãos, frutas, vegetais) distribuídas em diferentes países no final da década de 90 (Brown e Spillane, 1999). No Brasil existem três importantes coleções nucleares de cultivos, mantidas pela Embrapa. A coleção nuclear de milho, constituída por 300 subamostras usando os critérios morfológicos e de origem geográfica, considerando somente dados disponíveis de passaporte e de caracterização, sem a necessidade de um trabalho experimental adicional, que teria sido de custo elevado (Abadie et al., 2000; Cordeiro e Abadie, 2007). A CN de arroz, constituída por cerca de 600 subamostras baseadas em critérios morfológicos (sistemas de cultivo) e geográficos para estratificar a coleção (Abadie et al., 2005; Cordeiro e Abadie, 2007). E a CN de mandioca estabelecida a partir de critérios morfológicos (sistema de cultivo) e ecogeográficos (Cordeiro e Abadie, 2007). Mais recentemente, foi constituída a coleção nuclear de pupunha da Amazônia brasileira, o número de subamostras escolhidas baseou-se na representatividade e na distribuição geográfica das raças primitivas e populações, com seleção definitiva baseada na divergência genética entre subamostras com marcadores RAPD (Araújo, 2008).

Uma vez estabelecida a CN, é importante avaliar a sua adequação, ou seja, avaliar a sua representatividade e o seu grau de repetitividade. Este processo de validar a CN envolve usualmente compará-la com a CI da qual foi desenvolvida. O

conceito estatístico de representatividade é difícil de ser aplicado para definir a qualidade da CN em relação a CI, pois se apenas o significado estatístico for levado em conta, representatividade da CN pode significar média, variância, simetria e curtose semelhantes ou iguais à coleção base (Oliveira, 2007).

A representatividade da CN em relação à CI para fins de conservação significa manutenção da variabilidade genética, ou seja, são desejáveis igualdade entre médias, amplitudes, correlações entre as características das coleções. Significa também diferenças entre as distribuições uma vez que ocorrem desigualdades de variância (variância da CI < variância da CN) e curtose (curtose da CI > curtose da CN) (Oliveira, 2007).

Para Van Hintum et al. (2000), uma avaliação preliminar pode ser realizada por comparações de médias, amplitudes, frequências e variâncias para caracteres específicos entre os diferentes grupos da CN e a CI. Espera-se que os intervalos permaneçam similares enquanto que as médias movam-se em direção à mediana e as variâncias aumentem na CN, quando comparado com a CI.

Um erro muito frequente, no processo de validação, é usar somente a variância (Charmet e Balfourier, 1995) ou a média de características quantitativas como uma medida de validação. Esse erro é cometido quando o pesquisador baseia-se na hipótese de que a variância e/ou a média da CN deveriam ser o mais próximo possível daquelas da CI. Uma CN deve captar o maior número de alelos diferentes possíveis, incluindo os extremos, de forma que a distribuição das características quantitativas seja diferente e mais uniforme que na CI. Tal fato, certamente aumentará a variância e poderá alterar a média (Oliveira, 2007). Por esta razão a amplitude das características quantitativas tem sido o melhor critério para avaliar a qualidade da CN (Holbrook et al., 1993), indicando se todos os extremos estão representados (Holbrook, 1999).

Assim como nos estudos de diversidade, as coleções nucleares propostas levam em consideração apenas informações de passaporte ou de grupos isolados de características, sejam quantitativas, multicategóricas ou moleculares. Uma coleção nuclear por definição deve preservar o máximo de variabilidade genética. Assim, espera-se que ao reunir um maior número de dados de caracterização,

integrando avaliações de caracteres quantitativos, qualitativos, moleculares e fitopatológicos, as coleções estabelecidas sejam mais eficientes, principalmente no aspecto da conservação da variabilidade genética.

3. OBJETIVOS ESPECÍFICOS

- Realizar análise de diversidade de 67 subamostras de tomateiro do Banco de Germoplasma da UFV quanto aos caracteres quantitativos, fitopatológicos, multicategóricos e moleculares.
- Integrar os dados quantitativos, fitopatológicos, multicategóricos e moleculares para a análise da diversidade genética.
- Estabelecer subcoleções nucleares a partir de cada um dos conjuntos de dados, quantitativos, multicategóricos ou moleculares, e uma subcoleção que contemple a informação de todos esses dados.
- Comparar as diferentes subcoleções nucleares quanto a representatividade das mesmas.

4. MATERIAL E MÉTODOS

O estudo foi realizado a partir de um conjunto de dados de 67 subamostras de tomateiro do BGH-UFV (Tabela 1) caracterizadas quanto à 27 caracteres quantitativos (Marim, 2007), 34 caracteres qualitativos (Marim et al., 2002), 144 locus ISSR (Aguilera, 2007), reação à *Alternaria solani* (Rodrigues et al., 2006a), *Pseudomonas syringae* pv. *tomato* (Soares et al., 2007) e ao begomovírus *Tomato yellow spot vírus* (ToYSV) (Aguilera, 2007).

4.1 Caracterização morfoagronômica quantitativa (Marim, 2007) e multicategórica (Marim et al., 2002)

Os dados foram obtidos a partir de três ensaios distintos que utilizaram como testemunhas duas cultivares comerciais: Santa Clara (variedade de linha pura) e Débora Plus (variedade híbrida). O delineamento utilizado foi blocos ao acaso com três repetições e seis plantas por parcela.

Foram avaliadas características das fases de plântula, vegetativa e reprodutiva, bem como características agrônômicas e de qualidade dos frutos de acordo com as recomendações do International Plant Genetic Resources Institute – IPGRI (1996).

Tabela 1. Dados de origem de 67 subamostras de tomateiro do Banco de Germoplasma de hortaliças da UFV (BGH-UFV)

Subamostra	Origem	Subamostra	Origem	Subamostra	Origem
166	Não registrada	1019	Belo Horizonte-MG	1993	University of Purdue-USA
181	Não registrada	1020	Belo Horizonte-MG	2119	University of Purdue-USA
279	Não registrada	1211	Rio Casca-MG	2202	University of Purdue-USA
322	Não registrada	1214	São Gonçalo-MT	2203	University of Purdue-USA
349	Não registrada	1254	Não registrada	2205	University of Purdue-USA
468	Não registrada	1282	Londrina-PR	2208	University of Purdue-USA
489	Não registrada	1485	Pelotas (IPES)-RS	2211	University of Purdue-USA
773	Porto Simão-MT	1490	São Paulo (CAC)-SP	2213	University of Purdue-USA
850	Colatina-ES	1497	São Paulo (CAC)-SP	2214	University of Purdue-USA
970	Campinas-SP	1498	São Paulo (CAC)-SP	2216	University of Purdue-USA
975	Campinas-SP	1499	São Paulo (CAC)-SP	2219	University of Purdue-USA
978	Campinas-SP	1532	Belo Horizonte-MG	2223	University of Purdue-USA
980	Campinas-SP	1538	Anápolis-GO	2229	University of Purdue-USA
981	Campinas-SP	1706	São Paulo (CAC)-SP	2234	University of Purdue-USA
985	Campinas-SP	1708	São Paulo (CAC)-SP	3472	North Caroline-USA
987	Campinas-SP	1985	University of Purdue-USA	4006	Ipanema-MG
989	Campinas-SP	1987	University of Purdue-USA	4035	Não registrada
990	Campinas-SP	1988	University of Purdue-USA	4053	Santa Tereza-ES
991	Campinas-SP	1989	University of Purdue-USA	4054	Santa Tereza-ES
992	Campinas-SP	1990	University of Purdue-USA	4055	Colatina-ES
993	Campinas-SP	1991	University of Purdue-USA	4206	Não registrada
994	Campinas-SP	1992	University of Purdue-USA	4309	Ubá-MG
997	Campinas-SP				

4.1.1 Caracteres quantitativos avaliados

Na fase de plântula as características avaliadas foram: a) comprimento do cotilédone (CC), em mm; b) largura do cotilédone (LC), em mm; c) comprimento do hipocótilo (CompH), em mm e d) diâmetro do hipocótilo (DH), em mm.

Para as características vegetativas foram realizadas aleatoriamente duas medições por repetição, na folha ou entrenó imediatamente acima do terceiro cacho, sendo: a) comprimento da folha (CFo), em cm, medido da base da folha até a extremidade distal do último folíolo; b) largura da folha (LFo), em cm, medido no segundo par de folíolos; c) espessura do pecíolo principal (EPP) em mm, medido no pecíolo principal na posição mediana entre o segundo e terceiro par de folíolos; d) comprimento do entrenó (CE) em mm; e) diâmetro do entrenó (DE) em mm, medido na posição mediana do entrenó.

Na fase reprodutiva foram realizadas seis medições por repetição, sendo um fruto de cada planta da parcela, colhido do segundo ou terceiro cacho, quais sejam: a) comprimento do fruto (CFr), em mm; b) largura do fruto (LFr), em mm; c) espessura do mesocarpo (EM), em mm; d) espessura do endocarpo (EE), em mm; e) largura do eixo central (LEC), em mm; f) número de lóculos (NL). As características EM, EE e LEC foram medidas após o corte da seção mediana dos frutos. Para todas as características avaliadas as medições foram tomadas na posição do fruto que resultava na maior medida.

As características agronômicas avaliadas foram: a) número de frutos bons (NFB) em frutos.planta⁻¹; b) peso de frutos bons (PFB), em g.planta⁻¹; c) número de frutos ruins (NFR), em frutos.planta⁻¹; d) peso de frutos ruins (PFR), em g.planta⁻¹; e) peso médio de frutos (PMF), em g.fruto⁻¹; f) número total de frutos (NTF), em, frutos.planta⁻¹; g) peso total de frutos (bons + ruins) (PTF), em g.planta⁻¹; h) índice de precocidade (IP) em porcentagem, obtido pela razão entre a soma dos pesos de todos os frutos produzidos nas duas primeiras colheitas e o peso total de frutos (PTF), multiplicado por 100.

Considerou-se como frutos ruins aqueles com defeitos causados por pragas, doenças ou deficiências nutricionais. Frutos com rachaduras e lóculo

aberto foram considerados bons, sendo estas características avaliadas separadamente.

As características dos frutos foram obtidas a partir de medições realizadas em amostras de três frutos por parcela, sendo: a) acidez total (AT), o pH foi avaliado em acordo com a metodologia de Pregolato e Pregolato (1985); b) sólidos solúveis totais (SST), em °brix, medido com o auxílio de um refratômetro portátil; c) acidez total titulável (ATT), expresso em porcentagem de ácido cítrico; d) qualidade organoléptica (QO), obtida pela razão entre SST e ATT. Utilizaram-se frutos do segundo e terceiro cacho para as medições de qualidade de fruto.

4.1.2 Caracteres multicategóricos avaliados

Na fase de plântula foram avaliadas: a) cor do hipocótilo (CH): 1-verde, 2- $\frac{1}{4}$ lilás a partir da base, 3- $\frac{1}{2}$ lilás a partir da base, 4-lilás; b) Intensidade da cor do hipocótilo (ICH): 3-baixa, 5-intermediária, 7-alta; c) pilosidade no hipocótilo (PH): 0-ausente, 1-presente.

As características avaliadas na fase vegetativa foram: a) tipo de crescimento da planta (TCP): observados depois do início da colheita, sendo: 1-anão, 2-determinado, 3-semideterminado, 4-indeterminado; b) densidade de pilosidade na haste (DPH): 3-escasso, 5-intermediário, 7-denso, c) densidade de folhagem (DF): 3-escasso, 5-intermediário, 7-denso; d) atitude da folha (AtF): 3-semi-ereta, 5-horizontal, 7-prostrada; e) tipo da folha (TF): 1-anão, 2-tipo batata, 3-normal, 4-Peruvianum, 5-Pimpinelifolium, 6-Hirsutum, 7-outra;

Na fase reprodutiva avaliaram-se: a) tipo de inflorescência (TI): 1-geralmente unipariada, 2-ambos (parte uni e parte multipariada), 3-multipariada; b) cor da corola (CCor): 1-branca, 2-amarela, 3-laranja, 4-outra; c) tipo de florescimento da corola (TFC): 1-fechada, 2-aberta, d) coloração externa do fruto imaturo (CEFI): 1-branco esverdeado, 3-verde claro, 5-verde, 7-verde escuro, 9-verde muito escuro; e) presença de ombros verdes no fruto (POVF): 0-ausente, 1-presente; f) Intensidade dos ombros verdes no fruto (IOVF): 3-leve, 5-intermediário, 7-forte; g) pilosidade no fruto (PiIF): 3-escasso, 5-intermediário, 7-

denso; h) formato predominante do fruto (FPF): 1-achatado, 2-ligeiramente achatado, 3-arredondado, 4-muito arredondado, 5-forma de coração, 6-cilíndrico, 7-forma de pirâmide, 8-elipsóide, 9-outra; i) tamanho do fruto (TFr): 1-muito pequeno, 2-pequeno, 3-intermediário, 4-grande, 5-muito grande; j) homogeneidade do tamanho do fruto (HTF): avaliado dentro de uma mesma planta, sendo 3-baixa, 5-intermediária, 7-alta; k) coloração externa do fruto maduro (CEFM): 1-verde, 2-amarelo, 3-laranja, 4-rosa, 5-vermelho, 6-outra; l) intensidade da coloração externa do fruto maduro (ICEFM): 3-claro, 5-intermediário, 7-escuro; m) formato secundário do fruto (FSF): 1-achatado, 2-ligeiramente achatado, 3-arredondado, 4-muito arredondado, 5-forma de coração, 6-cilíndrico, 7-forma de pirâmide, 8-elipsóide, 9-outra; n) formato do ombro do fruto (FOF): 1-chato, 2-ligeiramente deprimido, 3-moderadamente deprimido, 4-fortemente deprimido; o) medida da área de cortiça ao redor da cicatriz do pedicelo (MACRCP): 3-pequena, 5-intermediária, 7-grande; p) facilidade de destacar o epicarpo (FDE): 3-fácil, 5-intermediário, 7-difícil; q) cor do epicarpo (CorE): a cor foi observada a partir do epicarpo destacado do fruto e analisando-o contra a luz, sendo 1-incolor, 2-amarela; r) cor do mesocarpo (CM): 1-verde, 2-amarelo, 3-laranja, 4-rosa, 5-vermelho, 6-outra; s) intensidade da cor do mesocarpo (ICM): 3-claro, 5-intermediário, 7-escuro; t) formato da sessão transversal do fruto (FSTF): 1-redondo, 2-angular, 3-irregular; u) formato da cicatriz estilar (FCE): 1-pontilhado, 2-estrelado, 3-linear, 4-irregular; v) forma da extremidade distal do fruto (FEDF): 1-recuado, 2-chato, 3-apontado; x) condição da cicatriz estilar (CCE): 1-aberto, 2-fechado; w) cor do eixo central (CEC): 1-verde, 2-branco, 3-transparente, 5-intermediário, 7-escuro; y) rachadura radial (RR): 1-linhas com cortiça, 3-leves, 5-intermediárias, 7-severas; z) rachadura concêntrica (RC): 1-linhas com cortiça, 3-leves, 5-intermediárias, 7-severas.

4.2 Caracterização fitopatológica

4.2.1 Reação a *Alternaria solani* (Rodrigues et al., 2006)

O experimento foi conduzido em dois ensaios, o primeiro teve início em novembro de 2005, e o segundo em fevereiro de 2007, ambos foram conduzidos na Horta Velha - UFV, em delineamento de blocos ao acaso, com três repetições e três plantas por parcela. Os procedimentos laboratoriais foram realizados no Laboratório de Epidemiologia de Departamento de Fitopatologia da Universidade Federal de Viçosa. O preparo do inóculo foi realizado a partir de isolados de *A. solani* das cidades de Domingos Martins (ES), Ponta Grossa (PR), Conselheiro Lafaiete (MG), Colméia (TO), Bueno Brandão (MG). Micélios do fungo foram retirados com água e pincel de cerdas finas, e em seguida induzidos à esporulação. Para induzir a esporulação, as placas de Petri foram submetidas à temperatura ideal para o fungo (25° C) e fotoperíodo de 12 horas de luz negra. Após três dias, tendo observado a esporulação, adicionou-se água destilada esterilizada a cada placa e, com o auxílio de um pincel de cerdas finas, foi realizado o desalojamento de conídios. Preparou-se a suspensão de conídios ajustada para a concentração de 1×10^3 conídios/mL. A suspensão foi colocada em pulverizador costal e procedeu-se a inoculação das subamostras, 60 dias após o transplântio.

As avaliações, após a inoculação das subamostras, foram realizadas a cada 3 dias. A característica avaliada foi área foliar lesionada pelo fungo *A. solani* (AFLAs), para tanto plantas receberam notas em função da porcentagem de área foliar lesionada, segundo o programa Severity Pro 1.0.

4.2.2 Reação a *Pseudomonas syringae* pv. *tomato* (Soares et al., 2007)

O experimento foi realizado em quatro ensaios diferentes, dois em janeiro de 2007, um em fevereiro de 2007, e o último em abril de 2007 sendo conduzidos na Horta Velha - UFV, em delineamento inteiramente ao acaso com 8 repetições. Os procedimentos laboratoriais foram realizados no Laboratório de Bacteriologia do Departamento de Fitopatologia da Universidade Federal de Viçosa. O isolado de *Pseudomonas syringae* pv. *tomato* foi coletado de plantas de tomateiro da cidade de Coimbra (MG) e mantido em geladeira em meio de cultura, o inóculo foi preparado no momento do uso, sendo ajustada para densidade óptica (DO) 540 = 0,1, que corresponde aproximadamente a 10^8 ufc de suspensão bacteriana para cada mL de solução salina 0,05%.

As mudas foram conduzidas até 4 ou 5 folhas definitivas (cerca de 15 cm de altura), sendo transplantadas para vasos de 300 mL, em casa de vegetação. Passados 10 dias do transplante, as mudas foram levadas à câmara de inoculação, onde inicialmente se aclimataram por 24 horas com o objetivo de promover a abertura dos estômatos. Após inoculação, as plantas permaneceram mais 24 horas na câmara para garantir a penetração da bactéria inoculada. Em seguida, as mudas foram levadas para uma câmara de crescimento com temperatura de 24°C ($\pm 2^\circ\text{C}$), onde permaneceram nesta condição por quinze dias. No final deste período, contou-se o número de pústulas em todas as folhas de cada planta.

4.2.3 Reação a begomovírus *Tomato yellow spot vírus* (ToYSV) (Aguilera, 2007)

Os experimentos foram conduzidos no Departamento de Fitopatologia da UFV, em casa de vegetação, em delineamento inteiramente casualizado, durante o período de Março/2006 e Abril/2007. Inicialmente plantas no estágio de quatro pares de folhas completamente expandidas foram agroinoculadas, empregando-se a estirpe LBA4404 de *Agrobacterium tumefaciens* transformada com os clones correspondentes ao DNA-A e -B do ToYSV.

A presença ou ausência do vírus foi avaliada de forma visual, classificando-se a planta como suscetível sempre que observado pelo menos um dos sintomas característicos do ToYSV, como mosaico severo, epinastia, manchas cloróticas e encarquilhamento foliar. A confirmação da infecção viral foi realizada por meio da PCR utilizando-se primers universais para o gênero begomovírus. A fenotipagem das subamostras foi realizada baseando-se na porcentagem de plantas infectadas (número de plantas positivas no teste de PCR em relação ao total de plantas testadas) dividida em cinco categorias: altamente resistentes (AR) 0-10%, resistentes (R) (11-20%), moderadamente resistente (MR) (21-40%), suscetíveis (S) (41-60%) e altamente suscetíveis (AS) (61-100%)

4.3 Caracterização molecular (Aguilera, 2007)

Folhas de três plantas por subamostra foram coletadas e congeladas a -80° C. A extração do DNA foi realizada em bulk de três folhas. As reações de amplificação foram realizadas em um volume final de 15 μ L contendo Tris-HCl 10 mM (pH 8.3), KCl 50 mM, $MgCl_2$ 2.4 mM, 0,25 μ M de cada um dos desoxinucleotídeos (dATP, dGTP, dCTP e dTTP), 0,5 μ M de *primer*, 0.6 unidades da enzima *Taq* DNA polimerase, 0,5% de BSA (albumina de soro bovino) e 20 ng de DNA. As amplificações foram realizadas num termociclador “GeneAmp PCR System 9600” (Applied Biosystem 9600), programado para uma etapa inicial de desnaturação de 4 min a 94° C, seguida de 35 ciclos de desnaturação (1 min a 4° C), pareamento (1 min a temperatura de 51° C) e extensão (2 min a 72° C). Por fim, uma fase de extensão final de 7 min a 72° C.

Dez *primers* ISSRs, foram selecionados para a reação de amplificação. Os produtos da amplificação foram visualizados em gel de agarose 1,2% corados com brometo de etídio a 0.02 μ L/ mL sob luz ultravioleta. As imagens foram documentadas por um sistema de fotodigitalização Eagle-Eye II (Stratagene). No total 144 bandas foram amplificadas. A interpretação dos padrões de bandas foi realizada pela atribuição de 1 à presença e 0 à ausência da banda para cada posição relativa.

4.4 Diversidade Genética

4.4.1 Diversidade genética de subamostras de tomateiro do BGH-UFV quanto aos caracteres quantitativos e fitopatológicos

Para avaliação da diversidade genética entre as subamostras, quanto aos caracteres quantitativos, realizou-se, inicialmente, a análise de variância agrupada pelo esquema 1 (Tabela 2), conforme descrito por Cruz et al.(2006), na qual foram identificadas as características com interação entre testemunhas e ensaios de efeito significativo, sendo estas descartadas. Para as características restantes efetuou-se a correção dos dados.

Tabela 2. Esquema da análise agrupada de variância para vários ensaios, nos quais são avaliados tratamentos comuns (testemunhas) e não comuns (genótipos)

FV	GL	SQ
Blocos/Ensaio	$(r-1)e$	SQB
Ensaio (E)	$e-1$	SQE
Testemunhas (Te)	$t-1$	SQTe
Interação (E x Te)	$(e-1)(t-1)$	SQTexE
Genótipos (G)/Ensaio	$\sum_{k=1}^e g_k - e$	SQG
(Te versus G)/E	E	SQGrupo
Resíduo	$(r-1) \left(\sum_{k=1}^e g_k - et - e \right)$	SQR
Total	$\left(ret + r \sum_{k=1}^e g_k \right) - 1$	

A correção dos dados foi realizada segundo Gomes (1990), a partir do efeito ambiental, estimado pela média das testemunhas em cada experimento subtraída da média geral destas em todos os experimentos, calculado para cada característica. As médias das subamostras para cada característica e experimento foram subtraídas do valor dos respectivos efeitos ambientais.

Paralelamente, realizou-se a análise de variância agrupada para os dados de reação a *Alternaria solani* e *Pseudomonas syringae*. A correção dos dados foi realizada a partir do efeito ambiental, estimado pelos dados da testemunha Santa Clara, da mesma forma procedida para os caracteres quantitativos.

Foi utilizado o teste de Dunnett, a 5% de probabilidade, na comparação da média de cada subamostra com a média das testemunhas para os caracteres quantitativos e fitopatológicos.

Na avaliação da diversidade genética das 67 subamostras foi empregado o agrupamento de Tocher, utilizando-se como medida de dissimilaridade a distância euclidiana média padronizada.

4.4.2 Diversidade genética de subamostras de tomateiro do BGH-UFV quanto aos caracteres multicategóricos

Inicialmente identificou-se as características para as quais não havia variação entre as subamostras, sendo a informação dessas descartadas. A correção dos dados também foi realizada conforme Gomes (1990). Após a correção procedeu-se a análise de diversidade utilizando-se como medida de dissimilaridade o complemento aritmético do índice de coincidência simples e agrupamento pelo método de Tocher.

Ressalta-se que os dados de caracterização quanto a reação ao begomovírus *Tomato yellow spot vírus* (ToYSV) foram analisados em conjunto com os caracteres multicategóricos.

4.4.3 Diversidade genética de subamostras de tomateiro do BGH-UFV quanto aos marcadores moleculares

Para a avaliação da diversidade genética em nível molecular, inicialmente foram identificadas as marcas polimórficas, sendo as demais descartadas. A análise da diversidade dos dados moleculares foi realizada por agrupamento pelo método de Tocher. Neste caso, foi utilizado o complemento aritmético do coeficiente de similaridade de Jaccard, como medida de distância.

A análise de diversidade também foi realizada utilizando-se como medida de dissimilaridade o complemento aritmético do índice de coincidência simples. A comparação da dissimilaridade observada pelos dois métodos foi realizada pelo teste Z de Mantel. Este teste permite comparar duas matrizes entre si, elemento por elemento, fornecendo um valor de correlação que quantifica o grau de relacionamento entre as matrizes, sendo testada a significância de r por meio de permutações (Manly, 1997) e do teste t, de Student, a 5% de probabilidade.

4.4.4 Comparação entre as matrizes de dissimilaridade obtidas dos caracteres quantitativos, multicategóricos e moleculares

Para verificar a existência de relação da diversidade entre cada um dos conjuntos de variáveis, os testes de t e Z de Mantel foram utilizados na comparação das matrizes de dissimilaridade obtidas a partir dos dados de diferentes naturezas.

4.4.5 Integração dos dados quantitativos, multicategóricos e moleculares

Duas estratégias foram utilizadas para a integração dos dados. A primeira foi a conversão dos dados quantitativos e fitopatológicos em multicategóricos, visando a obtenção de uma única matriz de dissimilaridade que contemplasse todos os caracteres independentes de sua natureza. A segunda estratégia

utilizada foi a obtenção das matrizes de dissimilaridade individualmente para cada conjunto de caracteres e em seguida a soma algébrica das mesmas.

4.4.5.1 Conversão dos dados quantitativos e fitopatológicos em multicategóricos

As médias dos dados quantitativos e fitopatológicos foram codificadas em classes a partir de sete subestratégias:

- Número de classes (K) pré-definido e divisão equitativa da amplitude (DEA): nesta estratégia as médias foram agrupadas em 2, 3, 4, 6, 8 e 10 classes. Os limites de cada classe foram estabelecidos a partir da divisão equitativa da amplitude.
- Número de classes (K) pré-definido em função do percentual equitativo dos dados (PED): nesta estratégia as médias foram agrupadas em 2, 3, 4, 6, 8 ou 10 classes. Cada classe foi estabelecida de forma que todas tivessem a mesma quantidade de dados.
- Número de classes (K) estimado pela raiz quadrada (RQ): nesta estratégia o número de classes foi estabelecido a partir do valor da raiz quadrada do número de subamostras avaliadas ($n=67$), resultando em 8 classes.
- Número de classes (K) estimado pela regra de Sturges (RS): nesta estratégia o número de classes foi estabelecido a partir da equação $K = 1 + 3,3 \log_{10} n$, em que n é o tamanho da amostra, resultando em 7 classes.
- Número de classes (K) definido em função da distribuição normal dos dados (DN): nesta estratégia os dados foram agrupados em seis classes que foram definidas com base na média e desvio-padrão (DP). O seguinte critério foi adotado: classe 1 - [mínimo, média - 2DP]; classe 2 - (média - 2DP, média-DP]; classe 3 - (média - DP, média]; classe 4 - (média , média + DP]; classe 5 - (média + DP, média + 2DP]; classe 6 - (média + 2DP, máximo].
- Número de classes (K) estabelecido a partir do teste de média Scott Knott (SK): nesta estratégia o número de classes foi estabelecido de acordo com o número de grupos formados para cada característica no teste de média

por Scott Knott, sendo assim, as subamostras foram classificadas para variáveis DH, EPP e ATT em 2 classes; CC, CE, PFB e reação à *Alternaria* em 3 classes; QO em 4 classes; CFo, LFr, EE, NL, NFB, NTF e SST em 5 classes; PTF e IP em 6 classes; LEC e reação à *Pseudomonas* em 7 classes; CFr e PMF em 8 classes.

- Número de classes (K) estabelecido conforme Scott (1979) (S): nesta estratégia o número de classes foi estabelecido a partir da equação $K = (A \cdot n^{1/3}) / (3,49 \cdot DP)$, em que A é a amplitude total, DP o desvio padrão de cada característica e n o número total de amostras, sendo assim, as subamostras foram classificadas para variáveis reação à *Alternaria* em 3 classes; EE e LEC em 4 classes; EPP, LFr, NL, PFB, PMF e PTF em 5 classes; DH, CC, CE, NFB, NTF, IP, SST, ATT, QO e reação à *Pseudomonas*; CFo e CFr em 7 classes.

As conversões de dados quantitativos em multicategóricos foram realizadas com o auxílio do programa Genes (Cruz, 2006). Após as conversões, novas matrizes de dissimilaridade foram obtidas, para cada subestratégia de codificação, utilizando-se como medida de dissimilaridade o complemento aritmético do índice de coincidência simples. Essas matrizes foram comparadas à matriz original por meio do teste Z de Mantel para se estabelecer a estratégia de conversão mais adequada.

Uma vez estabelecida a melhor subestratégia de conversão dos dados quantitativos e fitopatológicos em multicategóricos, realizou-se o estudo da diversidade entre as subamostras considerando-se todos os caracteres em uma única análise de diversidade. Para esta análise, utilizou-se o método de agrupamento de Tocher, a partir do complemento aritmético do índice de coincidência simples na obtenção da matriz de dissimilaridade (CONV).

4.4.5.2 Soma algébrica de matrizes

Após a obtenção das matrizes de dissimilaridade, individualmente para cada conjunto de caracteres, as mesmas foram somadas algebricamente, tomando-se o cuidado de realizar a padronização dos dados, obtendo-se outra matriz de dissimilaridade (SOMA), a qual foi comparada com a matriz CONV por meio do teste Z de Mantel. Também se realizou a análise de diversidade pelo método de agrupamento de Tocher a partir da matriz de dissimilaridade SOMA.

Cabe ressaltar que as matrizes de dissimilaridade obtidas individualmente, basearam-se na distância euclidiana média padronizada, complemento aritmético do índice de coincidência simples e complemento aritmético do índice de Jaccard, para os caracteres quantitativos, multicategóricos e moleculares, respectivamente.

4.4.6 Coleções Nucleares

4.4.6.1 Separação em grupos ou estratos geneticamente divergentes

As subamostras foram previamente separadas ou estratificadas quanto ao grupo comercial: Saladinha, Santa Cruz, Italiano ou Saladete, Maçã ou Caqui e Cereja. Essa separação foi realizada a partir da análise do formato de fruto, que foi avaliado por meio de notas (IPGRI, 1996) e das peculiaridades de cada grupo (Marim, 2007).

No estrato saladinha foram consideradas as subamostras com formato achatado a ligeiramente achatado; no estrato Santa Cruz considerou-se as subamostras com formato redondo a muito redondo; no grupo Italiano as subamostras com formato cilíndrico a elíptico. Adicionalmente, para serem consideradas dos grupos Saladinha e Santa Cruz, o peso médio dos frutos deveria variar entre 30g e 250g. Foram classificados como do grupo Maçã as subamostras com peso médio dos frutos superior a 250g e do grupo Cereja as subamostras com peso médio inferior a 30g.

Não foi utilizada a estratégia de separação por origem, pois parte dos acessos estudados não tiveram sua origem registrada.

4.4.6.2 Intensidade de amostragem e número de subamostras amostradas por estrato

Duas intensidades de amostragem das 67 subamostras foram avaliadas: 20 e 30%, representando 14 e 20 subamostras, respectivamente. Para a decisão do número de entradas selecionadas por estrato utilizou-se a estratégia logarítmica, em que o número de subamostras amostradas de cada estrato é calculado pela seguinte expressão (Oliveira, 2007):

$$NAE = \frac{\log a_i}{\sum_{i=1}^{ne} \log a_i} . nt$$

NAE = número de subamostras amostradas por estrato;

a_i = número de subamostras do i -ésimo estrato;

nt = número total de subamostras amostradas, definido pela intensidade de amostragem.

ne = número de estratos

4.4.6.3 Amostragem ou escolha das subamostras por estrato

A escolha das subamostras para compor cada uma das subcoleções foi realizada a partir da análise de diversidade dentro de cada estrato. Para tanto, matrizes de dissimilaridade entre as subamostras de cada estrato foram obtidas, a partir dos dados quantitativos e fitopatológicos, ou multicategóricos ou moleculares ou integrados por meio da conversão ou integrados por meio da soma de matrizes.

Para cada uma das matrizes de dissimilaridade de cada estrato identificou-se o valor máximo de distância (d_{max}) e este foi convertido em valor de similaridade utilizando-se a equação (Oliveira, 2007):

$$s = 100 - \left(\frac{100 \cdot d}{d_{\max}} \right)$$

s = similaridade,

d = é o valor de distância entre os indivíduos i e i' na matriz de dissimilaridade,

d_{\max} = valor máximo de distância.

Assim, a distância máxima foi igualada a zero. Como a matriz de dissimilaridade passou a ser de similaridade, ao realizar o agrupamento das subamostras pelo método de Tocher este saiu invertido, ou seja, foram agrupados os genótipos mais divergentes (Oliveira, 2007). A seleção dos acessos que compuseram as subcoleções foi realizada de acordo com a sequência do agrupamento pelo método de Tocher até que o número de subamostras atingisse o número pré-determinado de seleção para cada estrato de acordo com a intensidade de amostragem (Vasconcelos et al., 2007).

4.4.6.4 Subcoleções avaliadas

Foram avaliadas 10 subcoleções representativas das subamostras de tomateiro do BGH-UFV. Essas subcoleções foram definidas pela combinação entre a natureza dos dados avaliados e intensidade de amostragem, sendo:

- QUANT-20: amostragem multivariada a partir de dados quantitativos e intensidade de amostragem de 20%;
- QUANT-30: amostragem multivariada a partir de dados quantitativos e intensidade de amostragem de 30%;
- MULT-20: amostragem multivariada a partir de dados multicategóricos e intensidade de amostragem de 20%;
- MULT-30: amostragem multivariada a partir de dados multicategóricos e intensidade de amostragem de 30%;
- MOL-20: amostragem multivariada a partir de dados moleculares e intensidade de amostragem de 20%;

- MOL-30: amostragem multivariada a partir de dados moleculares e intensidade de amostragem de 30%;
- CONV-20: amostragem multivariada a partir da integração de dados quantitativos, multicategóricos e moleculares por meio da codificação de dados e intensidade de amostragem de 20%;
- CONV-30: amostragem multivariada a partir da integração de dados quantitativos, multicategóricos e moleculares por meio da codificação de dados e intensidade de amostragem de 30%;
- SOMA-20: amostragem multivariada a partir da integração de dados quantitativos, multicategóricos e moleculares por meio da soma de matrizes e intensidade de amostragem de 20%;
- SOMA-30: amostragem multivariada a partir da integração de dados quantitativos, multicategóricos e moleculares por meio da soma de matrizes e intensidade de amostragem de 30%;

4.4.6.5 Validação das subcoleções

O processo de validação das subcoleções envolveu a comparação das mesmas com a coleção inicial. Foram realizadas comparações levando-se em consideração o índice de coincidência da amplitude (CA) e a variância para cada grupo de característica, seja quantitativa, multicategórica ou molecular. Adicionalmente foi proposto um índice de retenção da variabilidade (IRV), para o qual as subcoleções também foram comparadas.

Comparação das amplitudes

O índice de coincidência da amplitude (CA) para cada subcoleção foi obtido para cada grupo de característica, seja quantitativa, multicategórica ou molecular, por meio da equação (Hu et al, 2000 e Wang et al, 2007):

$$CA = \frac{1}{n} \sum_{i=1}^n \frac{A_i SC}{A_i CI}$$

Em que:

CA = índice de coincidência da amplitude;

A_iSC = amplitude da i-ésima característica na subcoleção;

A_iCI = amplitude da i-ésima característica na coleção inicial;

n = número de características para determinado grupo.

Este índice também é conhecido por coeficiente de similaridade (Vasconcelos et al, 2007) ou índice de retenção da variabilidade – IRV (Diwan et al, 1995 e Oliveira, 2007). Ressalta-se, que até o momento, este índice vem sendo utilizado exclusivamente para a validação de coleções nucleares em relação à caracteres quantitativos. Em contrapartida, no presente estudo foram consideradas as amplitudes de todas as características, as quais foram utilizadas para a composição do índice de coincidência da amplitude para cada grupo de caracteres, sejam quantitativos, multicategóricos ou moleculares.

Comparação das variâncias

A variância foi estimada para cada uma das características, tanto na coleção inicial (CI) quanto nas subcoleções nucleares. A média das variâncias foi obtida para cada conjunto de características, seja quantitativa, multicategórica ou molecular, em cada uma das subcoleções e na CI.

Todas as comparações entre variâncias foram realizadas por meio do teste F (Snedecor e Cochran, 1980), dado por:

$$F_{\text{calculado}} = \frac{s_x^2}{s_y^2}$$

em que n_x e n_y são graus de liberdade associados a s_x^2 e s_y^2 , respectivamente. Foi adotada a maior variância como s_x^2 e a menor como s_y^2 , de modo a se obter sempre o valor de F calculado maior que 1. O valor de F calculado foi comparado com valor de F tabelado a 5% e 1% de probabilidade.

Análise gráfica da variabilidade

Para avaliar a representatividade das subcoleções nucleares quanto à frequência das classes de cada característica, sejam quantitativas, multicategóricas ou moleculares, os dados quantitativos, previamente codificados em multicategóricos, e os dados multicategóricos foram codificados em binários, ou seja, cada classe foi considerada como uma característica, e as subamostras receberam nota 1 quando pertenciam àquela classe e 0 caso contrário.

A frequência de subamostras que apresentavam cada uma das características foi estimada para todas as subcoleções nucleares. Característica presente em todos os indivíduos de uma subcoleção foi considerada fixada, enquanto foram consideradas características extintas aquelas com frequência nula nas subcoleções. Uma vez estimada a frequência de cada característica na coleção inicial e subcoleções os valores foram plotados no gráfico.

5. RESULTADOS E DISCUSSÃO

5.1 Diversidade genética de 67 subamostras de tomateiro do BGH-UFV quanto aos caracteres quantitativos e fitopatológicos

5.1.1 Análises univariadas

Pela análise de variância dos caracteres quantitativos verificou-se a existência de variabilidade entre as subamostras de tomateiro para todas as 27 características estudadas. Adicionalmente, observou-se interação significativa entre testemunhas e ensaios para 8 dessas, quais sejam: comprimento do hipocótilo (compH), largura do cotilédone (LC), largura de folha (LFo), diâmetro do entrenó (DE), espessura do mesocarpo (EM), número de frutos ruins (NFR), peso de frutos ruins (PFR) e acidez total (AT). Essas características foram descartadas das análises posteriores. Para reação a *Alternaria* sp. e *Pseudomonas* sp. também observou-se variabilidade entre as subamostras.

A comparação entre a média de cada subamostra com cada uma das testemunhas, pelo teste de Dunnett, permitiu identificar subamostras divergentes em relação às testemunhas que são variedades comerciais, para uma ou mais características quantitativas avaliadas (Tabela 3).

Considerando os caracteres da fase de plântula, observou-se para o comprimento do cotilédone (CC) que as subamostras 980 (30,09mm) e 1214 (32,04mm) apresentaram médias inferiores à das testemunhas, com valores médios de 47,08mm e 44,91mm, respectivamente. Para o diâmetro do hipocótilo

(DH) observou-se que as subamostras 970 e 2234 apresentaram médias inferiores apenas em relação à testemunha 'Santa Clara', com valores iguais à 1,66 e 1,51, respectivamente, enquanto a testemunha apresentou 2,43mm.

Para todas as características avaliadas na fase vegetativa verificou-se a ocorrência de subamostras diferentes das testemunhas. As subamostras 468, 1282, 1988, 1989, 1991, 1992 e 1993 apresentaram valores de comprimento de folha (CFo) entre 27,77 e 34,88cm. Esses valores foram inferiores aos observados para as testemunhas que apresentaram os respectivos valores de 45,89 e 46,56cm de comprimento foliar. Para a espessura do pecíolo principal (EPP), a subamostra 4309 (4,47mm) apresentou diferença significativa com relação só à testemunha 'Santa clara', com valor de 7,18mm. Já para o comprimento do entrenó (CE) a subamostra 1989 (13,64mm) apresentou valor inferior ao de ambas às testemunhas.

A obtenção de plantas mais compactas, ou seja, com menor espessura do pecíolo e menor comprimento da folha e do entrenó, é um dos objetivos dos programas de melhoramento do tomateiro destinados à indústria, uma vez que estas características são mais adequadas à colheita mecanizada (Carvalho et al., 2003). Adicionalmente, para o cultivo do tomateiro destinado ao consumo *in natura*, a redução do tamanho da planta possibilita a diminuição do investimento em mão de obra e insumos, principalmente defensivos (Campos et al., 1987; Oliveira et al., 1995).

Várias subamostras apresentaram diferenças significativas em relação às testemunhas para as características de fruto. Tal fato implica em grande diversidade de formatos e firmeza. Foram verificadas subamostras com dimensões tanto superiores quanto inferiores às testemunhas. Dentre essas, destacaram-se as subamostras 773, 980 e 1254 que apresentarem todas as medidas de fruto inferiores, o que permitiu caracterizá-las como tipicamente do grupo comercial cereja. Embora a maior parte do tomate comercializado *in natura* no país seja dos grupos Santa Cruz e Salada, outros grupos comerciais têm aumentado sua importância, principalmente os grupos Cereja e Italiano (Marim, 2007).

Tabela 3. Médias das avaliações nas fases de plântula, vegetativa, reprodutiva, dados agronômicos, qualidade dos frutos e reação a *Alternaria* e *Pseudomonas* de 67 subamostras de tomateiro e duas cultivares comerciais

Subamostra	DH ²		CC		CFo		EPP		CE		CFr		LFr	
166	2,39	ab ¹	41,68	ab	38,14	ab	5,79	ab	59,34	ab	52,09		63,12	ab
181	2,32	ab	40,79	ab	44,48	ab	6,71	ab	71,33	ab	47,56		60,08	ab
279	2,43	ab	43,74	ab	41,64	ab	5,84	ab	64,79	ab	44,29		45,02	
322	2,56	ab	49,06	ab	44,48	ab	7,02	ab	70,96	ab	41,92		39,35	
349	2,39	ab	44,99	ab	37,06	ab	6,59	ab	100,36	ab	49,06		64,22	ab
468	2,29	ab	35,91	ab	31,98		6,01	ab	66,74	ab	40,32		55,68	b
489	1,86	ab	35,01	ab	45,73	ab	7,84	ab	70,06	ab	34,72		44,42	
773	1,89	ab	33,94	b	37,64	ab	7,41	ab	51,99	ab	27,49		30,38	
850	1,89	ab	41,74	ab	42,23	ab	6,97	ab	87,86	ab	41,82		41,05	
970	1,66	b	40,24	ab	41,73	ab	6,32	ab	89,81	ab	61,32	b	53,92	b
975	1,98	ab	43,68	ab	45,31	ab	6,44	ab	74,41	ab	62,12	b	53,55	
978	2,08	ab	38,83	ab	44,14	ab	6,39	ab	48,73	ab	64,56	ab	67,28	ab
980	2,03	ab	30,09		38,06	ab	6,04	ab	42,58	ab	29,36		29,92	
981	2,26	ab	45,56	ab	44,48	ab	5,29	ab	72,76	ab	64,92	ab	72,95	ab
985	2,26	ab	40,44	ab	41,56	ab	6,92	ab	58,88	ab	40,46		45,42	
987	2,06	ab	45,09	ab	41,39	ab	6,51	ab	62,15	ab	54,59		51,92	
989	2,29	ab	43,19	ab	55,14	ab	6,54	ab	72,66	ab	65,29	ab	73,98	ab
990	2,33	ab	40,79	ab	40,06	ab	7,22	ab	61,06	ab	48,79		50,72	
991	2,26	ab	43,16	ab	42,98	ab	6,02	ab	79,81	ab	61,49	b	55,82	b
992	2,19	ab	47,86	ab	42,39	ab	7,37	ab	51,34	ab	57,19		43,22	
993	2,16	ab	49,99	ab	45,39	ab	7,67	ab	91,01	ab	55,76		50,65	
994	2,29	ab	43,59	ab	45,14	ab	6,84	ab	71,63	ab	60,12	b	50,52	
997	2,19	ab	44,39	ab	50,14	ab	5,89	ab	79,08	ab	63,49	ab	69,25	ab
1019	2,39	ab	45,49	ab	37,81	ab	4,74	ab	93,11	ab	52,79		68,35	ab
1020	2,63	ab	40,79	ab	42,48	ab	6,21	ab	83,96	ab	60,99	b	49,22	
1211	1,96	ab	52,76	ab	43,14	ab	6,16	ab	91,69	ab	58,62	b	52,15	
1214	2,29	ab	32,04		42,54	ab	5,47	ab	53,81	ab	55,66		76,28	ab
1254	2,16	ab	32,59	b	37,81	ab	6,12	ab	53,96	ab	31,24		34,35	
1282	2,23	ab	42,09	ab	27,77		5,40	ab	36,78	ab	52,49		61,04	ab
1485	2,49	ab	55,99	ab	39,72	ab	6,63	ab	54,16	ab	56,39		71,71	ab
1490	2,36	ab	53,59	ab	43,07	ab	6,98	ab	30,31	ab	51,43		57,46	ab
1497	2,63	ab	52,63	ab	38,97	ab	6,40	ab	56,74	ab	41,36		42,49	
1498	2,83	ab	59,46	a	43,42	ab	6,96	ab	48,64	ab	54,54		55,91	b
1499	2,86	a	50,59	ab	38,40	ab	6,73	ab	58,53	ab	58,47	b	63,53	ab
1532	2,36	ab	43,73	ab	39,05	ab	6,60	ab	51,83	ab	65,18	ab	54,50	b
1538	2,63	ab	55,59	ab	36,47	ab	6,58	ab	57,76	ab	60,98	b	56,99	b
1706	2,83	ab	55,06	ab	42,27	ab	7,56	ab	63,96	ab	49,42		53,40	
1708	2,69	ab	50,93	ab	42,13	ab	6,70	ab	28,38	ab	50,48		50,05	
1985	2,73	ab	42,09	ab	47,30	ab	6,00	ab	19,73	a	59,63	b	80,51	a
1987	2,53	ab	44,83	ab	41,07	ab	5,93	ab	45,03	ab	54,83		70,69	ab
1988	2,43	ab	46,13	ab	30,38		5,90	ab	45,81	ab	56,19		74,04	ab
1989	2,76	ab	42,69	ab	22,47		5,95	ab	13,64		48,07		58,34	ab
1990	2,39	ab	48,16	ab	53,22	ab	5,66	ab	48,33	ab	93,73		45,24	
1991	2,36	ab	42,76	ab	33,72		5,06	ab	44,61	ab	59,54	b	80,67	a
1992	2,86	a	47,73	ab	32,63		5,40	ab	34,63	ab	53,60		66,84	ab
1993	2,49	ab	42,73	ab	34,88		7,11	ab	46,34	ab	55,67		64,39	ab
2119	2,23	ab	46,03	ab	39,13	ab	5,86	ab	42,16	ab	57,68	b	75,84	ab
2202	3,08	a	49,94	ab	41,89	ab	6,53	ab	103,75	ab	60,38	b	77,37	ab
2203	1,98	ab	37,64	ab	38,73	ab	7,25	ab	64,97	ab	57,88	b	76,76	ab
2205	2,98	a	49,14	ab	46,31	ab	7,50	ab	62,48	ab	52,71		70,96	ab
2208	2,84	a	46,28	ab	49,06	ab	7,71	ab	60,78	ab	59,21	b	77,50	ab
2211	2,94	a	48,04	ab	45,64	ab	5,78	ab	107,20	b	63,01	b	74,26	ab
2213	2,54	ab	45,68	ab	49,14	ab	7,53	ab	66,53	ab	48,16		66,49	ab
2214	2,18	ab	45,91	ab	50,81	ab	6,46	ab	85,52	ab	54,43		71,91	ab
2216	2,68	ab	43,54	ab	44,81	ab	7,75	ab	57,58	ab	58,88	b	76,57	ab
2219	2,38	ab	43,51	ab	43,81	ab	6,50	ab	70,23	ab	57,92	b	72,03	ab
2223	2,58	ab	43,71	ab	47,23	ab	6,23	ab	69,70	ab	63,06	b	78,42	a

2229	2,54	ab	43,54	ab	42,31	ab	4,65	ab	59,83	ab	53,06		78,65	a
2234	1,51	b	35,18	ab	41,64	ab	6,70	ab	58,90	ab	62,05	b	85,42	
3472	2,14	ab	44,94	ab	43,48	ab	6,50	ab	73,63	ab	51,68		79,09	a
4006	2,31	ab	46,91	ab	37,64	ab	5,18	ab	49,78	ab	45,45		41,92	
4035	2,18	ab	47,04	ab	38,98	ab	4,95	ab	48,98	ab	39,04		40,44	
4053	2,24	ab	47,58	ab	44,14	ab	6,05	ab	73,17	ab	45,25		41,26	
4054	2,28	ab	47,91	ab	40,89	ab	7,41	ab	59,92	ab	85,15		56,11	b
4055	2,08	ab	47,91	ab	43,06	ab	6,35	ab	67,28	ab	51,21		49,29	
4206	2,51	ab	49,64	ab	40,23	ab	7,43	ab	59,32	ab	57,95	b	54,62	b
4309	2,51	ab	49,74	ab	35,89	a	4,46	b	65,35	ab	57,44		53,69	
Santa Clara	2,43	a	47,08	a	45,89	a	7,18	a	61,37	a	73,05	a	69,40	a
Débora	2,12	b	44,91	b	46,56	b	6,67	b	71,22	b	67,50	b	65,70	b

continuação Tabela 3

Subamostra	EE	LEC	NL	NFB	PFB	PMF	NTF						
166	48,26	ab	37,02	ab	3,92	ab	52,82	3666,03	ab	80,41		70,83	
181	45,76	ab	30,09	ab	4,58	ab	24,70	2994,72	ab	104,62	ab	40,04	ab
279	30,16		20,29		1,92	ab	46,80	3610,17	ab	81,37		65,52	a
322	28,82		23,82	b	2,25	ab	50,42	3905,08	ab	86,15		69,45	
349	50,89	ab	34,02	ab	3,92	ab	23,51	3100,11	ab	115,39	ab	34,54	ab
468	43,96	ab	36,75	ab	8,92		30,29	3070,17	ab	90,70		48,23	ab
489	35,16	b	20,45		2,25	ab	54,76	3052,83	ab	61,09		72,64	
773	21,79		9,72		1,92	ab	51,77	2404,72	ab	48,01		62,04	ab
850	28,99		16,39		3,58	ab	40,51	2819,78	ab	65,47		63,10	ab
970	36,39	b	27,82	ab	1,92	ab	36,10	3650,36	ab	101,85	ab	55,13	ab
975	36,62	b	26,22	ab	1,92	ab	32,68	3474,28	ab	104,97	ab	49,21	ab
978	55,22	ab	47,45	a	4,25	ab	19,80	3325,61	ab	167,16	ab	31,00	ab
980	23,46		9,55		2,25	ab	87,32	2907,50	ab	49,88		108,97	
981	62,56	a	52,99		6,92		16,76	2925,22	ab	169,02	b	26,72	b
985	35,29	b	19,42		1,92	ab	45,91	3134,39	ab	69,29		59,94	ab
987	40,26	ab	29,45	ab	1,92	ab	30,76	3219,42	ab	95,43	ab	48,51	ab
989	63,79		58,72		9,25		17,68	3072,89	ab	172,69		27,90	ab
990	39,19	ab	23,92	b	2,58	ab	43,62	3867,97	ab	91,68		55,99	ab
991	38,36	ab	22,55	b	1,92	ab	32,90	3334,17	ab	95,87	ab	52,88	ab
992	30,62		20,12		1,92	ab	34,34	3119,70	ab	83,26		54,15	ab
993	37,86	b	24,22	b	1,92	ab	40,35	3749,67	ab	95,15	ab	57,00	ab
994	36,86	b	19,12		2,25	ab	36,99	3658,95	ab	97,88	ab	62,12	ab
997	58,52	ab	58,02		6,58		18,70	3078,31	ab	156,47	ab	29,15	ab
1019	57,09	ab	44,75	a	5,58	ab	19,15	2742,89	ab	117,22	ab	29,68	ab
1020	33,22		28,12	ab	1,92	ab	34,39	3580,58	ab	102,38	ab	66,67	
1211	37,42	b	30,59	ab	2,25	ab	28,86	3364,97	ab	94,83	ab	46,38	ab
1214	68,82		60,25		12,25		17,21	3067,61	ab	183,66		29,23	ab
1254	27,66		19,52		3,58	ab	46,15	2505,92	ab	52,46		65,98	a
1282	54,10	ab	36,24	ab	5,93	ab	15,64	1913,06		115,28	ab	15,97	
1485	60,69	a	41,95	ab	4,86	ab	24,16	3454,67	ab	133,83	ab	35,74	ab
1490	45,09	ab	25,86	b	2,72	ab	47,88	3967,42	ab	80,02		67,40	
1497	33,10		14,52		2,08	ab	38,70	1691,25	ab	38,81		57,63	ab
1498	41,49	ab	22,57	b	2,27	ab	36,36	3255,22	ab	82,87		51,37	ab
1499	47,64	ab	31,38	ab	2,47	ab	22,51	2154,56	a	91,78		35,71	ab
1532	39,10	ab	22,65	b	2,14	ab	25,33	2281,67	a	84,51		43,94	ab
1538	41,58	ab	23,31	b	2,21	ab	27,80	2274,92	a	75,85		52,80	ab
1706	41,53	ab	19,99		2,28	ab	40,50	2829,89	ab	63,62		64,42	a
1708	36,33	b	15,67		2,29	ab	42,79	2889,50	ab	61,82		57,03	ab
1985	70,31		52,05		8,25		16,78	3173,67	ab	176,40		23,74	
1987	58,13	ab	39,47	ab	5,06	ab	10,61	1558,33		145,63	ab	21,71	
1988	67,81		46,27	a	8,53		12,28	1976,67		144,75	ab	16,77	
1989	49,85	ab	27,17	ab	4,00	ab	17,61	1455,39		78,40		20,32	
1990	32,78		15,30		2,25	ab	23,62	1842,39		73,94		45,17	ab
1991	73,08		54,66		13,51		15,28	3075,22	ab	185,24		21,22	
1992	58,33	ab	36,44	ab	4,92	ab	23,89	3079,17	ab	120,85	ab	27,11	b
1993	55,24	ab	37,88	ab	6,31	a	23,18	2256,00	a	91,25		25,40	b
2119	61,87	a	39,91	ab	5,29	ab	22,33	3550,00	ab	148,87	ab	36,66	ab
2202	67,02		51,28		7,00		15,08	3222,11	ab	190,33		22,88	

2203	66,55		52,39		7,19		9,13	a	1368,56		157,03	ab	14,87	
2205	60,49	a	47,83	a	8,50		17,44	ab	2205,89	a	128,73	ab	26,92	b
2208	67,84		51,19		7,22		11,08	ab	2449,72	ab	195,18		18,70	
2211	62,61	a	47,82	a	5,44	ab	12,36	ab	2574,06	ab	187,16		19,80	
2213	57,55	ab	44,80	a	8,21		14,17	ab	2136,97	a	147,74	ab	23,63	
2214	57,19	ab	44,40	ab	5,56	ab	19,13	ab	2946,97	ab	145,93	ab	30,71	ab
2216	66,61		49,18	a	6,72		8,42	a	2455,61	ab	234,12		18,92	
2219	61,50	a	44,94	a	6,00	a	18,93	ab	2966,56	ab	147,46	ab	28,05	ab
2223	68,03		52,77		7,61		9,00	a	2455,42	ab	224,28		21,17	
2229	68,82		47,36	a	8,89		16,29	ab	1723,56		117,80	ab	27,33	ab
2234	72,12		56,89		10,86		7,91	a	2303,67	a	229,92		14,07	
3472	66,53		49,05	a	6,60		15,91	ab	1861,39		123,56	ab	25,88	b
4006	29,99		18,61		2,39	ab	39,40	ab	395,39		16,23		49,97	ab
4035	30,36		16,47		2,34	ab	49,29		534,81		10,42		58,30	ab
4053	30,67		15,60		2,83	ab	54,26		952,67		14,72		68,85	
4054	38,75	ab	22,12	b	2,08	ab	19,04	ab	484,72		52,93		31,02	ab
4055	34,73	b	20,10		2,06	ab	44,69		1671,78		35,94		56,40	ab
4206	37,37	b	20,36		2,06	ab	51,35		2308,92	a	40,81		62,19	ab
4309	36,53	b	21,66	b	2,17	ab	36,26	ab	1964,56		54,98		45,28	ab
Santa Clara	50,74	a	37,80	a	2,81	a	23,93	a	3297,78	a	130,67	a	46,75	a
Débora	46,10	b	32,84	b	2,36	b	25,42	b	3532,72	b	132,35	b	43,56	b

continuação Tabela 3

Subamostra	PTF		IP		SST	ATT	QO		Alter	Pseud.			
166	5400,51	ab	4,97	ab	5,28	0,35	ab	15,03	ab	76,15	a	115,56	a
181	4783,44	ab	5,28	ab	4,77	b		9,57	ab	80,18	a	115,56	a
279	5328,20	ab	4,96	ab	5,50			14,87	ab	62,69	a	103,94	a
322	5653,45	ab	8,65	ab	4,90	b		13,44	ab	100,84	a	107,44	a
349	4655,62	ab	11,74	ab	4,23	ab		11,41	ab	97,62	a	95,44	a
468	4935,43	ab	9,32	ab	4,70	ab		11,29	ab	105,49	a	149,06	a
489	4501,23	ab	12,77	ab	5,00	b		13,41	ab	96,93	a	115,56	a
773	3576,23		11,55	ab	4,30	ab		12,74	ab	98,89	a	10,19	
850	4479,23	ab	9,10	ab	5,87			16,07	a	99,16	a	16,81	
970	5575,87	ab	1,80	ab	4,73	b		12,87	ab	36,30		61,94	a
975	5152,23	ab	1,67	ab	4,77	b		13,05	ab	100,95	a	21,94	
978	5498,84	ab	1,68	ab	4,27	ab		10,76	ab	97,89	a	69,19	a
980	4205,38	ab	14,56	ab	3,47	ab		9,43	ab	55,25	a	-25,31	
981	5132,29	ab	2,58	ab	5,00	b		13,84	ab	103,24	a	114,94	a
985	4544,68	ab	6,33	ab	5,73			13,97	ab	38,19	a	13,31	
987	4958,76	ab	9,26	ab	4,40	ab		12,76	ab	40,33	a	-8,31	
989	5066,54	ab	12,60	ab	4,53	ab		15,79	ab	99,00	a	12,06	
990	5385,65	ab	9,46	ab	4,20	ab		11,09	ab	53,20	a	-15,69	
991	5244,34	ab	5,04	ab	4,48	ab		11,59	ab	99,71	a	-14,81	
992	4732,65	ab	9,51	ab	4,67	ab		12,59	ab	36,02		-13,06	
993	5455,93	ab	2,99	ab	4,77	b		15,92	ab	30,85		10,31	
994	5901,62	ab	5,50	ab	4,90	b		17,90		30,14		-14,31	
997	4985,29	ab	12,52	ab	4,77	b		11,47	ab	101,96	a	90,31	a
1019	4510,81	ab	5,31	ab	4,10	ab		12,25	ab	51,65	a	204,69	
1020	5608,84	ab	6,43	ab	4,43	ab		14,47	ab	98,90	a	204,94	
1211	5427,87	ab	2,83	ab	4,17	ab		14,66	ab	33,74		289,69	
1214	5346,95	ab	13,06	ab	4,23	ab		11,73	ab	98,77	a	135,94	a
1254	3857,95	b	12,02	ab	4,80	b		10,40	ab	28,46		150,44	a
1282	1810,27		54,36		2,53			10,09	ab	63,96	a	129,56	a
1485	5316,22	ab	18,37	a	3,53	ab		10,92	ab	38,98	a	49,19	
1490	5192,41	ab	9,98	ab	3,60	ab		10,44	ab	31,66		67,06	a
1497	2047,19		15,84	ab	5,10	b		14,74	ab	33,82		53,19	
1498	4499,22	ab	12,03	ab	3,67	ab		11,52	ab	37,81		55,81	
1499	2983,36		9,86	ab	4,13	ab		13,02	ab	28,18		62,81	a
1532	3445,50		2,31	ab	4,33	ab		12,25	ab	28,38		55,06	
1538	3691,44		9,43	ab	3,97	ab		11,48	ab	39,16	a	93,94	a
1706	3909,11	b	13,71	ab	5,57			15,38	ab	29,74		77,56	a
1708	3530,00		19,91		3,70	ab		13,04	ab	37,35		93,06	a
1985	4462,33	ab	15,01	ab	4,07	ab		13,14	ab	40,60	a	73,06	a

1987	2570,05		27,12		4,23	ab	0,38	ab	11,00	ab	40,61	a	127,31	a
1988	2507,77		36,58		2,83	a	0,36	ab	7,92	ab	48,45	a	65,19	a
1989	1453,39		47,57		2,70	a	0,31	ab	8,83	ab	51,57	a	36,19	
1990	2382,50		8,50	ab	3,93	ab	0,27	ab	14,92	ab	34,86		58,94	a
1991	3781,61	b	42,75		3,47	ab	0,38	ab	9,36	ab	51,24	a	52,56	
1992	3360,28		27,12		3,37	ab	0,36	ab	9,50	ab	43,05	a	62,94	a
1993	2280,80		24,28		5,13	b	0,25	a	20,48		73,41	a	71,94	a
2119	5430,55	ab	12,80	ab	3,47	ab	0,30	ab	11,79	ab	34,45		94,44	a
2202	3723,41		13,02	ab	3,57	ab	0,40	ab	9,11	ab	28,32		84,56	a
2203	1563,05		9,16	ab	4,47	ab	0,51		9,10	ab	98,61	a	108,44	a
2205	2462,02		21,65		4,13	ab	0,41	ab	10,30	ab	35,22		96,19	a
2208	2955,27		16,39	ab	5,10	b	0,43	ab	12,26	ab	35,10		68,69	a
2211	3027,83		8,84	ab	3,77	ab	0,40	ab	9,67	ab	37,08		160,69	
2213	2512,75		26,44		3,57	ab	0,37	ab	10,52	ab	82,72	a	127,56	a
2214	3782,91	b	11,25	ab	4,17	ab	0,38	ab	11,10	ab	39,34	a	132,69	a
2216	3463,44		10,48	ab	3,90	ab	0,38	ab	10,50	ab	38,73	a	123,69	a
2219	3464,72		11,08	ab	3,77	ab	0,37	ab	10,28	ab	36,60		114,19	a
2223	3781,19	b	18,06	a	4,03	ab	0,39	ab	10,39	ab	86,46	a	120,19	a
2229	2339,33		12,66	ab	3,97	ab	0,31	ab	12,69	ab	39,58	a	129,69	a
2234	2605,44		14,16	ab	3,97	ab	0,33	ab	12,58	ab	96,60	a	184,94	
3472	2350,08		10,16	ab	4,30	ab	0,41	ab	10,77	ab	33,84		149,56	a
4006	-0,28		33,50		4,83	b	0,37	ab	13,20	ab	34,65		111,44	a
4035	87,83		26,43		4,43	ab	0,42	ab	11,05	ab	33,10		119,31	a
4053	790,27		20,55		4,77	b	0,37	ab	13,04	ab	31,16		127,06	a
4054	481,33		8,74	ab	3,73	ab	0,35	ab	10,73	ab	26,04		187,56	
4055	1514,55		14,18	ab	4,63	ab	0,39	ab	12,10	ab	31,30		111,31	a
4206	2120,00		11,28	ab	4,13	ab	0,34	ab	12,18	ab	26,62		141,81	a
4309	1731,11		7,32	ab	4,50	ab	0,32	ab	14,19	ab	43,71	a	145,19	a
Santa Clara	5463,77	a	6,70	a	3,70	a	0,33	a	11,46	a	74,97	a	108,56	a
Débora	5150,83	b	4,91	b	4,23	b	0,39	b	11,07	b	-	-	-	-

¹Médias seguidas da letra a e b, na coluna, são estatisticamente iguais, em nível de 5 % de probabilidade pelo teste de Dunnett, às testemunhas Santa Clara e Débora Plus, respectivamente.

²DH: diâmetro do hipocótilo; CC: comprimento do cotilédone; CFo: comprimento da folha; EPP: espessura do pecíolo principal; CE: comprimento do entrenó; CFr: comprimento do fruto; LFr: largura do fruto. EE: espessura do endocarpo; LEC: largura do eixo central; NL: número de lóculos; NFB: número de frutos bons; PFB: peso de frutos bons; PMF: peso médio de frutos; NTF: número total de frutos. PTF: peso total de frutos; IP: índice de precocidade; SST: sólidos solúveis totais; ATT: acidez total titulável; QO: qualidade organoléptica; *Alter.*: reação à *Alternaria* e *Pseud.*: reação à *Pseudomonas*.

Em relação aos dados agronômicos algumas subamostras se destacaram por apresentarem médias superiores a das testemunhas 'Santa Clara' e 'Débora Plus'. A subamostra 980 com as maiores médias para número de frutos bons (NFB) e número total de frutos (NTF), as subamostras 989, 1214, 1985, 1991, 2202, 2208, 2211, 2216, 2223 e 2234 para peso médio dos frutos (PMF) e as subamostras 1282, 1708, 1987, 1988, 1989, 1991, 1992, 1993, 2213, 4006, 4035 e 4053 para índice de precocidade (IP).

Para sólidos solúveis totais (SST) as subamostras 166, 279, 850, 985 e 1706 apresentaram médias, entre 5,50°brix e 5,87°brix, significativamente diferentes das médias das testemunhas 'Santa Clara' e 'Débora Plus'. As demais subamostras, exceto a 1282 (2,53°brix), não apresentaram diferenças significativas quanto a quantidade total de sólidos solúveis em relação às testemunhas 'S.Clara' e 'D. Plus', que apresentaram médias iguais à 3,7°brix e 4,2°brix, respectivamente.

Para acidez total titulável (ATT) destacou-se apenas a subamostra 2203 com média igual à 0,51, significativamente superior às testemunhas 'Santa Clara' e 'Débora Plus' que apresentaram médias iguais à 0,33 e 0,39, respectivamente. Para qualidade organoléptica (QO) os destaques foram para as subamostras 994 e 1993, que apresentaram valores iguais à 17,90 e 20,48, superiores as testemunhas com valores médios iguais à 11,46 e 11,07.

Quanto à reação a *Alternaria*, observou-se a ocorrência de 27 subamostras mais tolerantes a esta doença que a testemunha 'Santa Clara', as quais apresentaram áreas lesionadas menores. Estas subamostras foram as de números, 970, 992, 993, 994, 1211, 1254, 1490, 1497, 1498, 1499, 1532, 1706, 1708, 1900, 2119, 2202, 2205, 2208, 2211, 2219, 3472, 4006, 4035, 4053, 4054, 4055 e 4206. As médias das notas para área foliar lesionada variaram entre 26,04 e 37,81, enquanto a testemunha apresentou média igual à 74,97.

Considerando a reação à *Pseudomonas*, se destacaram as subamostras 773, 850, 975, 980, 985, 987, 989, 990, 991, 992, 993, 994, 1485, 1497, 1498, 1532, 1989 e 1991, por apresentarem menor número de pústulas em relação à testemunha 'Santa Clara'. Essas subamostras apresentaram número médio de

pústulas entre 0 e 55,81, enquanto a testemunha apresentou média de 108,56 pústulas.

5.1.2 Análises multivariadas

Mediante a análise de agrupamento pelo método de otimização de Tocher as 67 subamostras foram separadas em 14 grupos. Karasawa et al. (2005) avaliaram 70 subamostras de tomateiro do Banco de Germoplasma da Universidade Estadual do Norte Fluminense – UENF sendo essas separadas inicialmente em apenas dois grupos. Estes resultados indicam que o BGH-UFV é uma excelente fonte de recursos genéticos para a cultura do tomateiro.

Ficaram alocados no primeiro grupo 26 subamostras e a cultivar comercial Santa Clara, o que representa 40% do total das subamostras caracterizadas (Tabela 4). Este resultado mostra que de maneira geral, as subamostras deste grupo apresentaram características similares à da cultivar comercial. O fato de existirem várias subamostras semelhantes a cultivar comercial é útil aos melhoristas, pois caso alguma dessas subamostras possua um fenótipo de interesse, ausente na cultivar comercial, sua inclusão nos programas de melhoramento não prejudicaria as demais características (Marim, 2007). Dentre as subamostras do grupo 1 pode-se destacar a 993, 1497, 1498 e 1532, as quais apresentaram médias inferiores à testemunha nas avaliações tanto para a reação à *Alternaria* quanto à *Pseudomonas*. Assim, essas subamostras apresentam potencial para serem incluídas em programas de melhoramento que visem resistência a tais doenças, sem, contudo, provocar grandes alterações nas demais características comerciais.

No segundo grupo ficaram alocadas 21 subamostras, dentre essas se destacou a de número 2119 que apresentou características similares à testemunha Santa Clara, divergindo apenas quanto à reação a *Alternaria*, a área foliar lesionada nessa subamostra foi igual a 34,45, inferior à da testemunha 'S. clara' que apresentou área de 74,97.

Tabela 4. Agrupamento de 67 subamostras de tomateiro do BGH-UFV pelo método de Otimização de Tocher a partir dos dados quantitativos e reação à *Alternaria* e *Pseudomonas*

<i>Grupos</i>	<i>Subamostras</i>
1	975, 991, 987, 990, 992, 970, 1532, 985, 279, 993, 994, 166, 322, 850, 1706, 1538, 4206, 4055, 1497, 1499, 1498, 1708, 1490, Santa Clara, 349, 1020, 181
2	2214, 2219, 3472, 2223, 2216, 2213, 2205, 1987, 2208, 2211, 2202, 997, 1485, 2229, 2219, 1985, 981, 978, 1019, 1992, 1214
3	4006, 4035, 4053, 4309
4	1988, 1991, 1282, 1989
5	489, 773, 1254
6	468
7	980
8	1211
9	1993
10	2203
11	4054
12	1990
13	989
14	2234

No terceiro grupo ficaram alocadas as subamostras 4006, 4035, 4053 e 4309. Esse grupo é caracterizado por subamostras com dimensões menores de fruto, não excedendo ao peso médio de 55g.

As subamostras 1988, 1991, 1282, 1989 formaram o quarto grupo. Essas subamostras se destacam com bons índices de precocidade (IP) variando entre 36,58 a 54,38, no entanto, apresentaram número total de frutos (NTF) inferior à testemunha Santa Clara (Tabela 3).

O quinto grupo foi formado por três subamostras, todas com dimensões de fruto (CFr, LFr, EE e LEC) e peso médio de fruto (PMF) inferiores à testemunha 'S. clara'. As subamostras 468, 980, 1211, 1993, 2203, 4054, 1990, 989 e 2234 formaram sozinhas os grupos 6, 7, 8, 9, 10, 11, 12, 13 e 14, respectivamente.

5.2 Diversidade genética entre subamostras de tomateiro do BGH-UFV quanto aos caracteres multicategóricos e reação ao begomovírus

Dentre as 35 características avaliadas, 4 não apresentaram variação entre as subamostras, são essas: pilosidade do hipocótilo (PH), tipo de inflorescência (TI), cor da corola (CCor) e pilosidade no fruto (PilF). Essas não foram consideradas na análise da diversidade.

Nas avaliações quanto ao tipo de crescimento da planta (TCP) destacaram-se as subamostras 1989 do tipo anão e as subamostras 1988 e 1992 do tipo determinado. As subamostras 1282, 1993, 2208 e 2213 foram classificadas como do tipo semideterminado. Todas as demais subamostras avaliadas, inclusive as testemunhas 'Santa Clara' e 'Débora Plus' apresentaram tipo de crescimento indeterminado.

Para a característica densidade da folhagem (DF), as subamostras 161, 181, 349, 468, 489, 981, 990, 991, 1019, 1282, 1497, 1988, 1990, 1991, 1992, 2005, 2213, 2229 e 2234 foram classificadas como escassas enquanto as testemunhas foram classificadas como intermediárias. A existência de plantas com tipo de crescimento determinado e densidade de folhagem escassa é interessante,

sobretudo quando objetiva-se o desenvolvimento de variedades mais aptas à colheita mecanizada.

Para as características relacionadas ao fruto uma ampla variabilidade foi observada tanto para a coloração, formato e tamanho. Para esta última, as subamostras 773 e 980 foram classificadas como muito pequenas, como era de se esperar de acordo com a classificação dessas no grupo comercial cereja.

Nas avaliações para resistência ao *begomovirus* destacaram-se as subamostras 773 e 1532, classificadas como moderadamente resistentes (MR), enquanto as testemunhas comerciais 'Santa Clara' e 'Débora Plus' foram classificadas como altamente suscetíveis (AS). Essas subamostras apresentam potencial para serem incluídas em programas que visem o aumento da resistência ao *begomovirus* nas cultivares comerciais.

Mediante a análise de agrupamento pelo método de otimização de Tocher, as 67 subamostras também foram separadas em 14 grupos (Tabela 5), o que evidenciou a presença de diversidade genética em relação aos caracteres multicategóricos avaliados.

No primeiro grupo foram alocados 25% das subamostras, todas com tipo de crescimento indeterminado (TCP), tipo de folha normal (TF), cor do epicarpo (CorE) amarela, tamanho de fruto (TFr) não excedendo a classe intermediária e altamente suscetíveis ao *begomovirus*.

O segundo grupo foi formado por 18% das subamostras, todas com TF normal e CorE amarelo, dentre as subamostras que compõem esse grupo destaca-se a 1532 por apresentar alta resistência ao *begomovirus*.

As subamostras que formaram o terceiro grupo apresentam parte do hipocótilo lilás (CH), presença de ombros verdes nos frutos (POVF) e fruto de tamanho intermediário, em comum. Apresentam, ainda, a origem americana, todos vindos da Universidade de Purdue.

O grupo 4 foi formado por subamostras com TCP indeterminado, e TFr variando entre pequeno e intermediário, e coloração externa quando maduros (CEFM) rosa avermelhada.

O quinto e sexto grupos foram formados por subamostras com tipo de folha normal, estes apresentam ainda baixas densidades de pilosidade na haste (DPH) e foliar, enquanto aqueles apresentam tamanho de fruto intermediário.

Tabela 5. Agrupamento de 67 subamostras de tomateiro do BGH-UFV pelo método de Otimização de Tocher a partir de dados multicategóricos e reação ao begomovírus

<i>Grupos</i>	<i>Subamostras</i>
1	993, 994, 970, 992, 975, 1020, 985, 279, 990, 1254, 991, 850, 322, 1211, 987, 166, 980
2	1497, 1990, 1708, 1706, 1498, 1706, 1499, 1490, 1538, 1532, 2119, 1992, 1485
3	2216, 2229, 2211, 2219, 2223, 2205, 2208, 2202
4	4006, 4053, 4035, 4206, 4055, 4054, 4309
5	981, 1019, 997, 349, 978, 989, 468, 1214
6	2213, 2234
7	489, 773
8	1985, 1991, 1988, 1282
9	2203, 3472
10	2214
11	1987
12	1993
13	181
14	1989

Frutos de menor tamanho são característicos de subamostras pertencentes ao grupo 7, essas são muito similares para a maior parte das características. Diferenças sutis entre essas subamostras podem ser observadas na densidade de folhagem (DF) e na reação ao begomovírus ToYSV, sendo a 773 altamente resistente a este.

O grupo 8 foi formado por 4 subamostras, todas apresentam frutos ligeiramente achatados. Já o nono grupo foi formado por duas subamostras que são muito similares, mais de 50% das características obtiveram notas coincidentes e as poucas diferenças encontradas são muito sutis, sugerindo a possibilidade de constituírem réplicas.

Os grupos 10, 11, 12, 13 e 14 foram formados por uma única subamostra cada.

5.3 Diversidade genética entre subamostras de tomateiro do BGH-UFV quanto aos caracteres moleculares

Apenas 52 das 144 bandas amplificadas foram polimórficas, representando 36,8% do total de locos amplificados. A baixa quantidade de locos polimórficos observados pode ser um reflexo do estreitamento de base genética ocorrido durante a domesticação da cultura.

Mediante a análise de agrupamento pelo método de otimização de Tocher os acessos foram separados em apenas 2 grupos. Um dos grupos foi formado apenas pela subamostra 980 e o outro pelos demais acessos avaliados (Tabela 6). Tal fato corrobora com a hipótese de estreitamento da base genética do tomateiro durante o processo de domesticação

Realizando um reagrupamento, excluindo-se a subamostra 980, 11 subgrupos dentro do grupo 1 foram formados, o que evidenciou os diferentes níveis de dissimilaridade dentro deste grupo.

As 67 subamostras também foram agrupadas, pelo método de Tocher, só que utilizando como medida de dissimilaridade o complemento aritmético do índice de coincidência simples. Neste caso, também obteve-se o mesmo agrupamento

das subamostras do que quando se utilizou o complemento aritmético do índice de Jaccard. A correlação entre as matrizes de dissimilaridade obtidas por estes dois índices de distâncias genéticas foi de 0,98, significativa a 5% de probabilidade pelo teste Z de Mantel.

Tabela 6. Agrupamento de 67 subamostras de tomateiro do BGH-UFV pelo método de Otimização de Tocher a partir de dados moleculares

<i>Grupos</i>	<i>Sub-Grupo</i>	<i>Subamostras</i>
1	1.1	970, 991, 987, 1211, 975, 279, 985, 349, 989, 1990, 978, 981, 2234, 1020, 1991, 1989, 322, 994, 850, 1214, 1254, 993, 1019, 2229, 4054, 4006, 166, 990, 489, 468, 1988, 992, 1498, 4053, 1532, 1485, 2214, 3472, 1538
	1.2	2205, 2208, 1708, 2203, 2202, 1992, 1985, 1706, 4035, 2019, 1499
	1.3	4055, 4309, 4206, 2211, 2213
	1.4	997, 2223
	1.5	181, 773
	1.6	1993, 2219
	1.7	2216
	1.8	1282
	1.9	1987
	1.10	1490
	1.11	1497
2		980

5.4 Comparação entre as análises de diversidade com base em caracteres quantitativos, multicategóricos e moleculares

Os agrupamentos realizados pelo método de Tocher, utilizando-se como medidas de dissimilaridade a distância euclidiana média padronizada (para os dados quantitativos) e a obtida a partir do complemento do índice de coincidência (para os dados multicategóricos) entre as subamostras permitiram a formação de número igual de grupos, no entanto a distribuição das subamostras entre os grupos foi diferenciada. No agrupamento pelo método de Tocher utilizando-se os dados quantitativos observou-se maior concentração de subamostras nos grupos 1 e 2. Já no agrupamento utilizando-se o conjunto de dados multicategóricos a distribuição das subamostras entre os 14 grupos foi de forma mais equitativa.

Em contrapartida, no agrupamento realizado pelo método de Tocher utilizando-se as informações moleculares observou-se que, embora exista a diversidade fenotípica, a diversidade molecular é mais restrita. Todas as 67 subamostras foram agrupadas em apenas dois grupos. Sendo o grupo 1 formado por 66 dessas.

Segundo Gomes (2007), a formação de número de grupos diferentes e a distribuição diferenciada das subamostras nos grupos, mesmo quando o número de grupos formados é igual, evidenciam uma discordância entre os agrupamentos realizados. Assim, pode-se concluir que neste estudo, a diversidade entre as subamostras com base em um conjunto de caracteres, seja quantitativo, multicategórico ou molecular não pode ser extrapolada para outro destes conjuntos.

Para cada par de matrizes de dissimilaridade, obtidas com base nos grupos de caracteres, sejam quantitativos, multicategóricos ou moleculares, foi estimado o coeficiente de correlação e este testado pelo teste Z de Mantel (Tabela 7). Estes valores de correlação, embora significativos, foram de baixa magnitude, corroborando que a diversidade a partir de um conjunto de dados não pode ser extrapolada para outro conjunto dados de natureza diferente. A significância de valores de correlação de baixa magnitude deve-se ao elevado número de graus de

liberdade, 2209 neste caso, para testar os valores de correlação pelo teste t e ao elevado número de observações (2211) comparadas nas matrizes de dissimilaridade para o teste Z de Mantel.

A decisão sobre qual grupo de variáveis utilizar para a determinação da divergência genética e de qual medida de dissimilaridade utilizar, depende do objetivo do estudo. A utilização de caracteres multicategóricos é prática, econômica e demanda menor tempo, quando comparada a caracteres quantitativos e moleculares (Marim et al., 2002). Em muitas culturas, apesar dos caracteres quantitativos serem de difícil mensuração, esses têm sido preferidos em estudos de diversidade, por apresentarem importância comercial (Gomes, 2007).

A observação de valores de correlação menores que 0,5 sugere que a diversidade genética obtida com base nos caracteres quantitativos não explica a diversidade dos caracteres multicategóricos nem a diversidade obtida por marcadores moleculares, pois segundo Tatineni et al. (1996) valores de r são estatisticamente significativos a 1% de probabilidade, se r é superior a 0,50 quando o número de unidades comparadas for maior do que 15.

Vários autores que se propuseram a comparar as distâncias de dissimilaridade de diferentes conjuntos de dados relataram resultados diversos.

Ao avaliar a correlação entre medidas de dissimilaridade utilizadas na determinação da divergência genética de mandioca para caracteres multicategóricos e quantitativos, Gomes (2007) observou valor de correlação igual a 0,09. O autor atribuiu a inexistência de correlação entre as duas medidas de dissimilaridade à diferença do controle genético para os diferentes tipos de caracteres estudados.

Oliveira (2005), ao comparar dissimilaridades genéticas e fenotípicas de açaizeiro, observou valores de correlação inferiores à 0,05 e de magnitude negativa. Koehler-Santos et al. (2003) também não encontraram concordância no padrão de variação em germoplasma de *citrus*, quando compararam as matrizes geradas com dados morfológicos e marcadores SSR.

Tabela 7. Correlação entre matrizes de dissimilaridades de caracteres quantitativos, multicategóricos e moleculares pelo teste Z de Mantel e teste t

<i>Dissimilaridades</i>	<i>Teste Z de Mantel</i>
Quantitativo com Multicategórico	0,46808** ++
Quantitativo com Molecular	0,20946**
Multicategórico com Molecular	0,21379** ++

** Significativo a 1% de probabilidade pelo teste t.

++ Significativo a 1% de probabilidade, pelo teste de Mantel baseado em 1000 simulações.

Vale ressaltar que marcadores moleculares baseiam-se em fragmentos aleatórios do DNA genômico. Logo a possibilidade dos fragmentos analisados estarem associados aos caracteres morfoagronômicos estudados é mínima, pois esses caracteres são quantitativos e devem envolver vários genes (Oliveira, 2005).

Resultado divergente foi encontrado por Alves (2002) que obteve $r=0,72$ quando avaliou a associação entre 21 locos microssatélites e 19 descritores botânicos-agronômicos em 31 subamostras de cupuaçuzeiro.

Segundo N'Goran et al. (1994), a possibilidade de concordância entre a dissimilaridade genética e fenotípica varia com o tamanho da amostra, sendo, geralmente, reduzida quando a amostra apresenta tamanho pequeno.

Levando em consideração que o teste Z de Mantel apresenta resultado consistente a partir de valores cofenéticos acima de 0,7 (Rohlf, 2000), deduz-se que conjuntos de informações obtidas a partir de dados de diferentes naturezas não estão plenamente correlacionados, não oferecendo, dessa forma, suporte para extrapolar os resultados de um conjunto de dados para outro (Oliveira, 2005). Isso implica que todos os três conjuntos de caracteres são importantes para a organização da diversidade do germoplasma do tomateiro.

O estabelecimento de coleções nucleares visa à conservação do máximo da variabilidade genética entre as amostras. Assim, é recomendável a consideração do máximo de informações de caracterização, sejam essas a partir de dados quantitativos, fitopatológicos, multicategóricos ou moleculares, uma vez que a diversidade observada para um ou outro conjunto de dados não pode ser extrapolada. Dessa forma, metodologias de integração dos dados de diferentes naturezas são necessárias para o estabelecimento de coleções nucleares mais efetivas, especialmente na preservação da variabilidade genética.

5.5 Integração de dados qualitativos, quantitativos, moleculares e fitopatológicos em estudos de diversidade do tomateiro

5.5.1 Conversão dos dados quantitativos em multicategóricos

No total, 17 matrizes de dissimilaridade, a partir da conversão dos dados quantitativos em multicategóricos, foram obtidas. As estimativas do coeficiente de correlação entre essas matrizes e a matriz de dissimilaridade obtida a partir dos dados originais são apresentadas na Tabela 8. Segundo Rohlf (2000), a correlação entre duas matrizes é consistente apenas a partir de valores cofenéticos acima de 0,7 pelo teste Z de Mantel. Assim, dentre as subestratégias de conversão dos dados quantitativos em multicategóricos, apenas as codificações pela distribuição eqüitativa da amplitude (DEA) em 2, 3 e 4 classes foram consistentes, com valores de correlação significativos a 1% de probabilidade, tanto pelo teste t quanto Z de Mantel. Destas destacou-se a conversão em três classes, cujo valor de correlação foi o mais elevado (0,782).

Para a subestratégia PED (percentual equitativo dos dados), em que foi mantido o mesmo número de subamostras por classe, observou-se que com o aumento do número de classes a correlação entre as matrizes de dissimilaridade diminuiu. Ademais, o maior valor de correlação (0,629), embora significativo, não superou 0,7. Assim, a conversão dos dados por esta subestratégia não se mostrou consistente.

Cabe ressaltar que as subestratégias RQ (raiz quadrada) e RS (regra de Sturges) definem apenas o número de classes, sendo a alocação dos valores nas classes realizada também pela distribuição equitativa da amplitude. Desta forma, os valores de correlação entre estas duas subestratégias, para o mesmo número de classes da DEA, são coincidentes (Tabela 8).

Na Figura 1 é apresentada a distribuição de freqüência dos valores de correlação para as subestratégias DEA e PED em função do número de classes. É evidente que existe um número ótimo de classes para a conversão dos dados,

uma vez que pela subestratégia mais adequada (DEA) a correlação foi máxima para três classes e diminuiu à medida que o número de classes aumentou.

Determinada a melhor subestratégia de codificação dos dados (DEA-3) procedeu-se o estudo de diversidade genética com a integração dos dados de diferentes naturezas. Assim, a análise de diversidade genética foi realizada considerando 104 caracteres (52 marcas moleculares, 31 características multicategóricas e 21 quantitativas). O agrupamento das 67 subamostras pelo método de Tocher, baseado em uma única matriz de dissimilaridade (CONV) obtida a partir do complemento aritmético do índice de coincidência simples é apresentado na Tabela 9.

Foram alocados no grupo 1 cerca de 46% das subamostras, todas de origem nacional. Destaca-se que o segundo grupo foi formado apenas por subamostras oriundas da Universidade de Purdue. Tal fato se repetiu na formação do grupo 4, com exceção da subamostra 1485, que foi alocada também nesse grupo e sua origem é de Pelotas (IPES-RS).

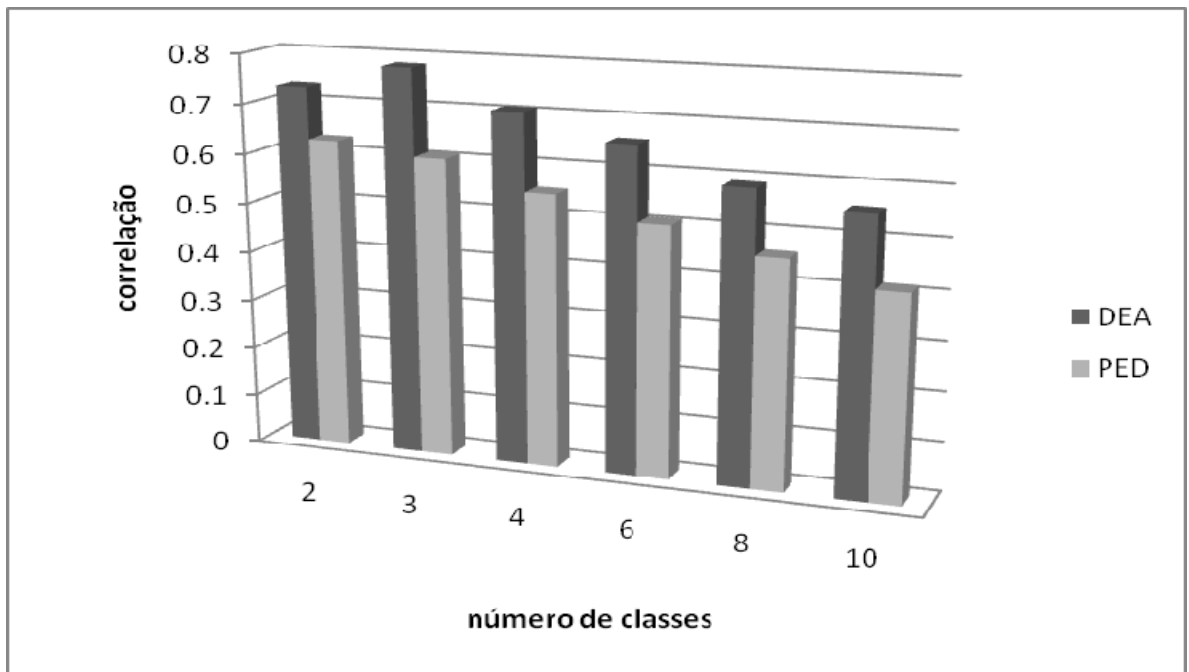


Figura 1. Correlação entre a matriz original obtida a partir de dados quantitativos e fitopatológicos com as matrizes obtidas a partir da codificação de dados pelas subestratégias DEA e PED

Tabela 8. Correlação entre a matriz de dissimilaridade a partir dos dados quantitativos originais e as matrizes de dissimilaridade obtidas por 17 estratégias de codificação destes dados

Subestratégia	Número de classes	Valor de r
DEA	2	0,732** ++
	3	0,782** ++
	4	0,707** ++
	6	0,658** ++
	8	0,590** ++
	10	0,557** ++
PED	2	0,629** ++
	3	0,607** ++
	4	0,552** ++
	6	0,508** ++
	8	0,461** ++
	10	0,415** ++
RQ	8	0,590** ++
RS	7	0,624** ++
DN	6	0,664** ++
SK	variou de acordo com as características	0,655** ++
S	variou de acordo com as características	0,678** ++

** Significativo a 1% de probabilidade pelo teste t.

++ Significativo a 1% de probabilidade, pelo teste Z de Mantel baseado em 1000 simulações.

Subestratégias: DEA - número de classes pré definido e divisão equitativa da amplitude; PED: número de classes pré definido e percentual equitativo dos dados; RQ: número de classes estimado pela raiz quadrada de n; RS: número de classes estimado pela regra de Sturges; DN: número de classes fixo pressupondo distribuição normal dos dados; SK: número de classes estabelecido a partir do teste de média Scott Knott; S: número de classes estabelecido conforme Scott, 1979

Tabela 9. Agrupamento de 67 subamostras de tomateiro do BGH-UFV pelo método de Otimização de Tocher a partir da integração de dados pela estratégia CONV e subestratégia DEA-3

<i>Grupos</i>	<i>Subamostras</i>
1	4006, 4053, 4035, 4055, 4206, 4309, 4054, 1532, 1538, 1498, 1708, 1706, 1499, 1990, 2229, 987, 1211, 985, 279, 970, 1020, 993, 166, 975, 991, 992, 994, 322, 990, 1254, 850
2	2205, 2208, 2216, 2219, 2211, 2223, 2202, 2214, 3472, 2234, 2213, 2203
3	981, 989, 978, 1019, 997, 349, 1214, 468
4	1485, 2119, 1992, 1985, 1991, 1988, 1987
5	489, 773
6	1282, 1989
7	1490, 1497
8	181
9	1993
10	980

5.5.2 Soma algébrica de matrizes

Após a soma algébrica das matrizes de dissimilaridade padronizadas, utilizou-se o agrupamento das 67 subamostras de tomateiro pelo método de Tocher, o que permitiu a formação de 8 grupos (Tabela 10). O primeiro grupo foi formado por 36% das subamostras, com a maioria de origem nacional, exceto as subamostras 166, 279, 322 e 349 que não possuem origem registrada.

No segundo grupo foram alocadas apenas subamostras de origem norte americana. No grupo três foram alocadas principalmente subamostras brasileiras, com exceção da subamostra 1990. O grupo quatro foi formado por subamostras em sua maioria norte americanas, com exceção das subamostras 1282 e 1485.

Ao comparar as duas estratégias de integração dos dados, CONV e SOMA, observou-se certa correspondência entre os agrupamentos, embora o número de grupos formados seja diferente, 10 para a estratégia CONV e 8 para SOMA (Tabelas 9 e 10). Essa aparente correspondência foi confirmada pelo alto valor da correlação, $r=0,96$, significativo a 5% de probabilidade pelo teste Z de Mantel, entre as matrizes obtidas pelas duas estratégias de integração. Entretanto, cabe ressaltar que o agrupamento a partir da integração dos dados pela estratégia CONV foi superior à SOMA, uma vez que esta estratégia permitiu uma maior discriminação das subamostras, com alguns grupos sendo subdivididos.

Tabela 10. Agrupamento de 67 subamostras de tomateiro do BGH-UFV pelo método de Otimização de Tocher a partir da integração de dados pela estratégia SOMA

<i>Grupos</i>	<i>Subamostras</i>
1	970, 975, 991, 987, 994, 993, 985, 279, 990, 1020, 322, 166, 850, 1211, 1254, 489, 349, 981, 978, 773, 1019, 989, 1532, 1498
2	2205, 2208, 2219, 2229, 2216, 2214, 2223, 2211, 2202, 2213, 3472, 2234, 2203, 2119
3	4055, 4206, 4309, 4053, 4006, 4035, 4054, 1538, 1708, 1706, 1499, 1497, 1990
4	1988, 1991, 1992, 1985, 1987, 1485, 1989, 1282
5	181, 468, 1214, 997
6	1490
7	1993
8	980

5.6 Coleções Nucleares

5.6.1 Separação em grupos ou estratos geneticamente divergentes

Das 67 subamostras avaliadas, 38 foram classificadas como pertencentes ao grupo comercial ou estrato Saladinha, 21 ao estrato Santa Cruz, 2 ao estrato Italiano, 3 ao estrato Maçã e 3 ao estrato Cereja (Tabela 11).

5.6.2 Número e subamostras amostradas por estrato

A quantidade de subamostras selecionadas dentro de cada estrato foi proporcional ao logaritmo de seu tamanho. Nas amostragens a 20% de intensidade, o número de subamostras amostradas nos estratos Saladinha, Santa Cruz, Italiano, Maçã e Cereja foram 5, 4, 1, 2 e 2, respectivamente. Enquanto nas amostragens a 30% de intensidade, o número de subamostras amostradas para os mesmos estratos foram 8, 6, 2, 2 e 2, respectivamente.

A estratégia logarítmica assegura que grupos contendo poucas subamostras sejam representados na subcoleção e, também, que aqueles com grande número de subamostras contribuam relativamente com menos subamostras na coleção nuclear (Oliveira, 2007). Assim, a amostragem logarítmica aumenta a probabilidade de capturar os alelos menos frequentes comparativamente à amostragem aleatória (Brown, 1989; Van Hintum, 1999). Ela evita a amostragem excessiva de subamostras de estratos grandes e aumenta o número de subamostras amostradas nos estratos pequenos, reduzindo o viés do tamanho do grupo (Oliveira, 2007).

Tabela 11. Classificação de 67 subamostras de tomateiro do Banco de Germoplasma de hortaliças da UFV (BGH-UFV) quanto ao grupo comercial

Subamostras	Grupo comercial	Subamostras	Grupo comercial	Subamostras	Grupo comercial	Subamostras	Grupo comercial
166	SC	990	SAL	1532	SAL	2211	SAL
181	SAL	991	SC	1538	SC	2213	SAL
279	SC	992	SC	1706	SC	2214	SAL
322	SAL	993	SC	1708	SC	2216	MA
349	SAL	994	SC	1985	SAL	2219	SAL
468	SAL	997	SAL	1987	SAL	2223	MA
489	SAL	1019	SAL	1988	SAL	2229	SAL
773	CER	1020	SC	1989	SAL	2234	MA
850	SAL	1211	SAL	1990	ITA	3472	SAL
970	SC	1214	SAL	1991	SAL	4006	SC
975	SC	1254	CER	1992	SAL	4035	SC
978	SAL	1282	SAL	1993	SAL	4053	SC
980	CER	1485	SAL	2119	SAL	4054	ITA
981	SAL	1490	SC	2202	SAL	4055	SC
985	SAL	1497	SC	2203	SAL	4206	SC
987	SAL	1498	SC	2205	SAL	4309	SC
989	SAL	1499	SAL	2208	SAL		

SAL: Saladinha; SC: Santa Cruz; ITA: Italiano ou Saladete; MA: Maçã ou Caqui; CER: Cereja.

A partir do agrupamento, para cada estrato, pelo método de Tocher invertido, baseado nas matrizes de dissimilaridade obtidas de dados de diferentes naturezas e considerando as intensidades de amostragens propostas (20 e 30%), foram estabelecidas as seguintes coleções nucleares:

- QUANT-20: subamostras 850, 1989, 2211, 989 e 1211 do estrato Saladinha; subamostras 994, 4053, 1020 e 1490 do estrato Santa Cruz; subamostra 4054 do estrato Italiano; subamostras 2216 e 2234 do estrato Maçã e subamostras 980 e 1254 do estrato Cereja.
- QUANT-30: subamostras 850, 1989, 2211, 989, 1211, 1991, 489 e 1282 do estrato Saladinha; subamostras 994, 4053, 1020, 1490, 4006 e 970 do estrato Santa Cruz; subamostras 1990 e 4054 do estrato Italiano; subamostras 2216 e 2234 do estrato Maçã e subamostras 980 e 1254 do estrato Cereja.
- MULT-20: subamostras 489, 1987, 2203, 989 e 2213 do estrato Saladinha; subamostras 975, 4053, 1708 e 991 do estrato Santa Cruz; subamostra 1990 do estrato Italiano; subamostras 2216 e 2234 do estrato Maçã e subamostras 980 e 1254 do estrato Cereja.
- MULT-30: subamostras 489, 1987, 2203, 989, 2213, 1989, 2202 e 181 do estrato Saladinha; subamostras 975, 4053, 1708, 991, 4035 e 1497 do estrato Santa Cruz; subamostra 1990 e 4054 do estrato Italiano; subamostras 2216 e 2234 do estrato Maçã e subamostras 980 e 1254 do estrato Cereja.
- MOL-20: subamostras 181, 2211, 1282, 1987 e 1499 do estrato Saladinha; subamostras 1490, 1498, 1497 e 4055 do estrato Santa Cruz; subamostra 4054 do estrato Italiano; subamostras 2216 e 2234 do estrato Maçã e subamostras 773 e 980 do estrato Cereja.
- MOL-30: subamostras 181, 2211, 1282, 1987, 1499, 997, 1985 e 1993 do estrato Saladinha; subamostras 1490, 1498, 1497, 4055, 4035 e 994 do estrato Santa Cruz; subamostra 1990 e 4054 do estrato Italiano; subamostras 2216 e 2234 do estrato Maçã e subamostras 773 e 980 do estrato Cereja.

- CONV-20: subamostras 489, 1993, 2202, 181 e 1987 do estrato Saladinha; subamostras 975, 4035, 1490 e 4309 do estrato Santa Cruz; subamostra 1990 do estrato Italiano; subamostras 2216 e 2234 do estrato Maçã e subamostras 980 e 1254 do estrato Cereja.
- CONV-30: subamostras 489, 1993, 2202, 181, 1987, 850, 2208 e 1282 do estrato Saladinha; subamostras 975, 4035, 1490, 4309, 991 e 1497 do estrato Santa Cruz; subamostra 1990 e 4054 do estrato Italiano; subamostras 2216 e 2234 do estrato Maçã e subamostras 980 e 1254 do estrato Cereja.
- SOMA-20: subamostras 850, 2202, 1282, 181 e 1993 do estrato Saladinha; subamostras 991, 4035, 1490 e 975 do estrato Santa Cruz; subamostra 1990 do estrato Italiano; subamostras 2216 e 2234 do estrato Maçã e subamostras 980 e 1254 do estrato Cereja.
- SOMA-30: subamostras 850, 2202, 1282, 181, 1993, 2208, 489 e 2211 do estrato Saladinha; subamostras 991, 4035, 1490, 975, 4053 e 166 do estrato Santa Cruz; subamostra 1990 e 4054 do estrato Italiano; subamostras 2216 e 2234 do estrato Maçã e subamostras 980 e 1254 do estrato Cereja.

As subcoleções estabelecidas são bem diferentes. Apenas as subamostras 980, 2216 e 2234, pertencentes aos estratos Cereja, Maçã e Maçã, respectivamente, estão presentes em todas as subcoleções formadas. Para os estratos com maior número de subamostras, Saladinha e Santa Cruz, não foram observadas subamostras comuns a todas as subcoleções. As diferenças observadas decorrem da utilização de dados de diferentes naturezas no estabelecimento das subcoleções. Isto sugere que subcoleções baseadas em grupos de caracteres, sejam quantitativas, qualitativas ou moleculares, isolados, não abrangem a diversidade genética como um todo. Cabe ressaltar que as diferenças entre as subcoleções, CONV e SOMA, são devidas aos distintos processos utilizados na integração dos dados.

5.6.3 Validação das subcoleções

Uma vez estabelecida a subcoleção, é importante avaliar a sua adequação, especialmente em relação a sua representatividade, ou seja, sua capacidade de manutenção da variabilidade existente na coleção inicial.

Dentre as subcoleções, considerando 20% de intensidade de amostragem, destacaram-se CONV-20 e SOMA-20, obtidas a partir da integração de dados, pois essas apresentaram valores de índice de coincidência da amplitude (CA) superiores a 0,80 em relação a todos os grupos de características, sejam quantitativos, multicategóricos ou moleculares (Tabela 12). Segundo Diwan et al. (1995), Hu et al. (2000) e Wang et al. (2007), uma subcoleção é representativa quando seu índice de coincidência da amplitude é de no mínimo 80%. Desta forma, as subcoleções estabelecidas a partir de cada conjunto de caracteres, QUANT-20, MULT-20 ou MOL-20, não foram consideradas representativas, pois apresentaram valores de CA de 0,79, 0,77 e 0,78 para os grupos de caracteres, moleculares, quantitativos e multicategóricos, respectivamente.

A variância da subcoleção para cada grupo de caracteres é outro parâmetro que deve ser levado em consideração na avaliação da representatividade das subcoleções. Mantendo a amplitude das características, espera-se aumento das variâncias nas subcoleções em relação à coleção inicial, uma vez que o número de indivíduos amostrados é menor. No entanto, observou-se a diminuição significativa da variância nas subcoleções MULT-20 e MOL-20 em relação ao grupo de caracteres quantitativos (Tabela 12), o que indica a não inclusão de indivíduos com valores fenotípicos extremos no processo de amostragem dessas subcoleções.

Para a subcoleção QUANT-20 observou-se aumento significativo da variância para o grupo de caracteres quantitativos. Como essa estratégia apresentou elevado valor de CA, pode-se inferir que esse aumento da variância se deve à maior frequência das classes extremas nessa subcoleção e, ainda, que as classes intermediárias foram pouco representadas. Assim, a avaliação do CA e da variância devem ser complementares na validação e escolha da melhor estratégia

de obtenção de subcoleções nucleares, possibilitando concluir que para subcoleções com um mesmo valor de CA, as de menor variância são mais representativas.

Considerando os valores de CA e variância das características quantitativas, as subcoleções CONV-20 e SOMA-20 foram as mais eficientes.

Para caracteres multicategóricos, em que as classes intermediárias também devem ser representadas nas subcoleções, as considerações em relação ao CA e alteração da variância são as mesmas acima relacionadas às características quantitativas. São desejáveis subcoleções com alto índice de coincidência da amplitude acompanhado por variâncias de variação não significativas em relação à coleção inicial.

Já em relação ao grupo de características moleculares, as coleções nucleares devem ser estabelecidas de forma a preservar o maior número de alelos e aumentar a frequência dos alelos raros (alelos com frequência menor que 5%). Uma coleção nuclear ideal é aquela que consegue manter para um mesmo loco, o maior número de alelos com igual frequência. Em se tratando de marcadores moleculares dominantes, como os ISSR, a subcoleção mais adequada é aquela em que as frequências dos indivíduos que possuem a presença ou a ausência da marca são as mesmas. Desta forma, uma subcoleção será mais eficiente quanto maior for sua variância.

Considerando os três grupos de caracteres, quantitativos, multicategóricos e moleculares, a estratégia CONV-20 se destacou em relação à SOMA, apresentando elevados valores de CA, estimativas menores e não significativas das variâncias para os grupos de caracteres quantitativos e multicategóricos e maior variância em relação aos caracteres moleculares.

Tabela 12. Índice de coincidência da amplitude (CA) e variância média das características nas subcoleções formadas a partir de 20% de intensidade de amostragem dos dados

Subcoleção ¹	Quantitativos ²	Multicategóricos	Moleculares
QUANT-20	0,90 (204325,8 ^{**}) ³	0,80 (1,6508 ^{ns})	0,79 (0,1106 [*])
MULT-20	0,77 (119051,5 [*])	0,82 (2,0951 ^{**})	0,81 (0,1135 ^{**})
MOL-20	0,85 (120873,5 [*])	0,78 (1,5764 ^{ns})	0,92 (0,1431 ^{**})
CONV-20	0,86 (138148,4 ^{ns})	0,85 (1,9171 ^{ns})	0,90 (0,1359 ^{**})
SOMA-20	0,91 (144725,7 ^{ns})	0,83 (1,9431 [*])	0,88 (0,1307 ^{**})
CI	(144671)	(1,6938)	(0,0987)

¹ QUANT-20: subcoleção formada a partir de dados quantitativos; MULT-20: subcoleção formada a partir de dados multicategóricos; MOL-20: subcoleção formada a partir de dados moleculares; CONV-20: subcoleção formada a partir da integração de dados quantitativos, multicategóricos e moleculares por meio da codificação de dados quantitativos em multicategóricos; SOMA-20: subcoleção formada a partir da integração de dados quantitativos, multicategóricos e moleculares por meio da soma algébrica de matrizes; CI: coleção inicial.

² Corresponde ao grupo de caracteres do qual foram estimados CA e variância.

³ Entre parênteses: valores de variância; * ** significativos a 5 e 1% de probabilidade, ^{ns} não significativo a 5% de probabilidade pelo teste do F máximo.

Em relação a intensidade amostragem igual a 30%, as subcoleções QUANT-30 e MULT-30 apresentaram valores de CA acima de 80% (Tabela 13). Entretanto, essas apresentaram estimativas de variâncias elevadas e significativas em relação aos grupos de caracteres quantitativos e multicategóricos, respectivamente, sendo consideradas menos eficientes.

A subcoleção MOL-30 se mostrou representativa, com valores de CA acima de 80% acompanhados de variâncias de magnitudes adequadas para os três grupos de caracteres. Esse resultado indica que, para essa intensidade de amostragem, apenas a caracterização molecular pode ser suficiente para o estabelecimento de subcoleções nucleares eficientes em manter toda a variabilidade do germoplasma.

Tabela 13. Índice de coincidência da amplitude (CA) e variância das subcoleções formadas a partir de 30% de intensidade de amostragem dos dados

Subcoleção	Quantitativos	Multicategóricos	Moleculares
QUANT-30 ¹	0,94 (213747,5 ^{**}) ²	0,85 (1,7624 ^{ns})	0,83 (0,1020 ^{ns})
MULT-30	0,89 (163845,9 ^{ns})	0,90 (1,9856 [*])	0,94 (0,1131 ^{**})
MOL-30	0,92 (158051,7 ^{ns})	0,85 (1,6108 ^{ns})	0,98 (0,1334 ^{**})
CONV-30	0,92 (150664,5 ^{ns})	0,86 (1,8708 ^{ns})	0,94 (0,1279 ^{**})
SOMA-30	0,91 (172489,8 [*])	0,85 (1,8032 ^{ns})	0,90 (0,1219 ^{**})
CI	(144671)	(1,6938)	(0,0987)

¹ QUANT-30: subcoleção formada a partir de dados quantitativos; MULT-30: subcoleção formada a partir de dados multicategóricos; MOL-30: subcoleção formada a partir de dados moleculares; CONV-30: subcoleção formada a partir da integração de dados quantitativos, multicategóricos e moleculares por meio da codificação de dados quantitativos em multicategóricos; SOMA-30: subcoleção formada a partir da integração de dados quantitativos, multicategóricos e moleculares por meio da soma algébrica de matrizes; CI: coleção inicial.

² Corresponde ao grupo de caracteres do qual foram estimados CA e variância.

³ Entre parênteses: valores de variância; * ** significativos a 5 e 1% de probabilidade, ^{ns} não significativo a 5% de probabilidade pelo teste do F máximo.

5.6.4 Análise gráfica da variabilidade das subcoleções

A codificação dos dados quantitativos e multicategóricos em binários, juntamente com os dados moleculares, resultou em um total de 288 classes, sendo 63 para os dados quantitativos, 173 para os multicategóricos e 52 para os moleculares.

Nas Figuras 2, 3 e 4 são representadas as frequências das classes, referentes às características quantitativas, multicategóricas e moleculares, respectivamente, para a subcoleção CONV-20. Em cada gráfico, para cada uma das classes de caracteres são representadas a frequência de subamostras pertencentes aquela classe tanto na coleção inicial quanto na subcoleção e o desvio entre as mesmas. Observa-se que apenas duas das classes quantitativas não foram mantidas na subcoleção (indicadas pelas setas), ou seja, nenhuma das subamostras escolhidas para compor CONV-20 apresentava tal característica.

Em relação aos caracteres multicategóricos, 142 das 173 classes foram representadas na subcoleção, ou seja, 82% dos fenótipos foram mantidos, pois estavam presentes em pelo menos uma das subamostras que compuseram CONV-20. Para os dados moleculares a variabilidade foi mantida para 92% dos locos, o que representa a perda de cinco alelos (indicados pelas setas) no processo de amostragem, seja através da fixação da presença da marca ou da fixação da ausência da marca.

Na Figura 5 são representadas as frequências das classes, referentes às características multicategóricas para a subcoleção MULT-20. Observa-se que, embora esta subcoleção apresente índice de coincidência de 0,82 (Tabela 12), apenas 77% da variabilidade foi mantida para este o conjunto de classes de caracteres, o que correspondeu à perda de 40 das 173 classes, ou seja, 40 das classes apresentaram frequência nula na subcoleção. Estes resultados ratificam que no estabelecimento de coleções nucleares as informações de variância devem ser complementares às do índice de coincidência da amplitude.

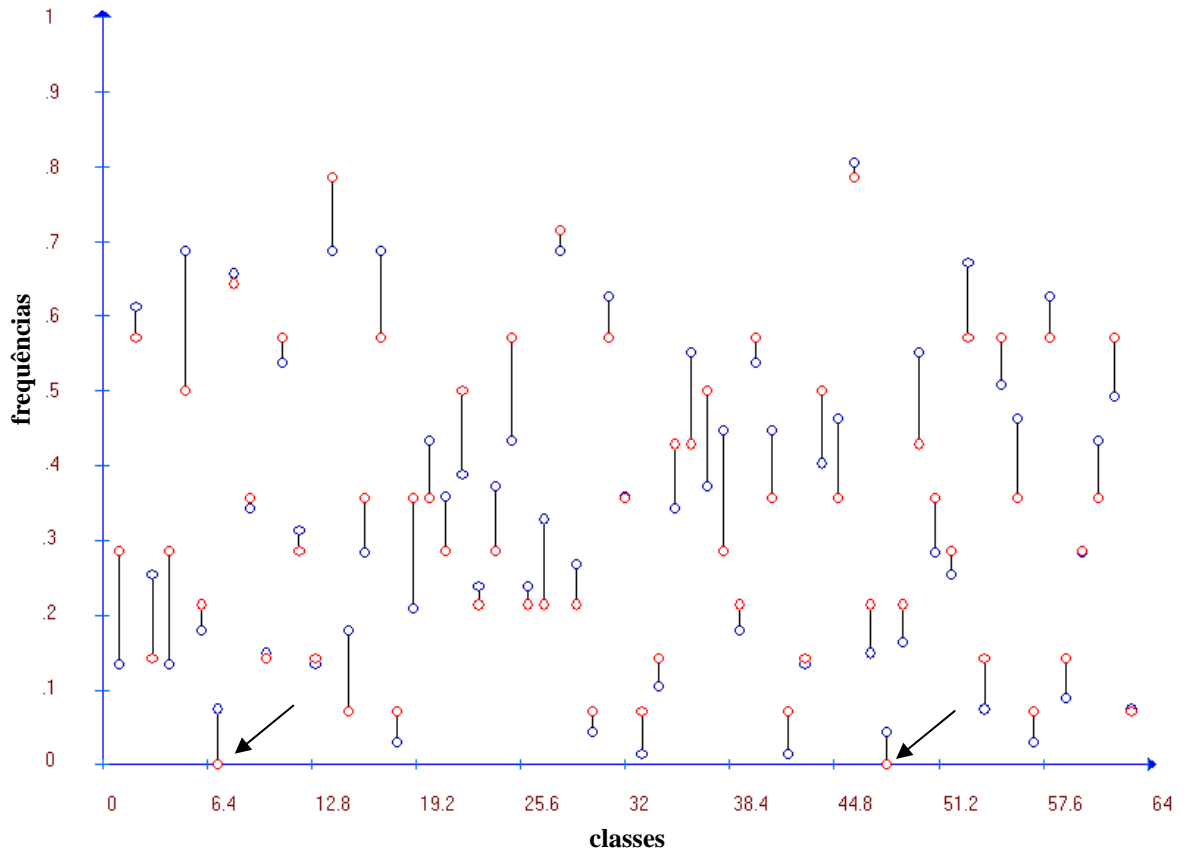


Figura 2. Variação de frequência das 63 classes dos caracteres quantitativos. ○ Frequência das classes na coleção inicial, ○ frequência das classes na subcoleção CONV-20. As setas indicam classes de caracteres que apresentaram frequência nula na subcoleção.

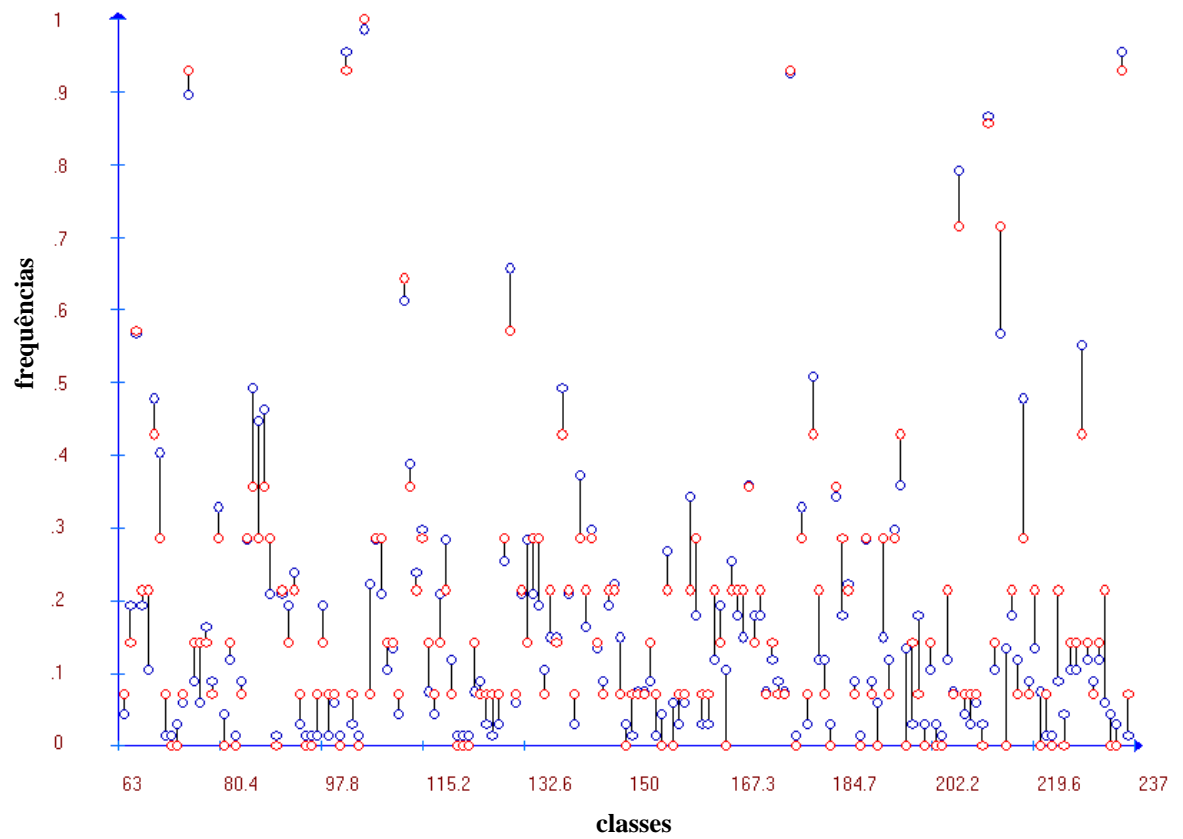


Figura 3. Variação de frequência das 173 classes dos caracteres multicategóricos. ○ Frequência das classes na coleção inicial, ○ frequência das classes na subcoleção CONV-20.

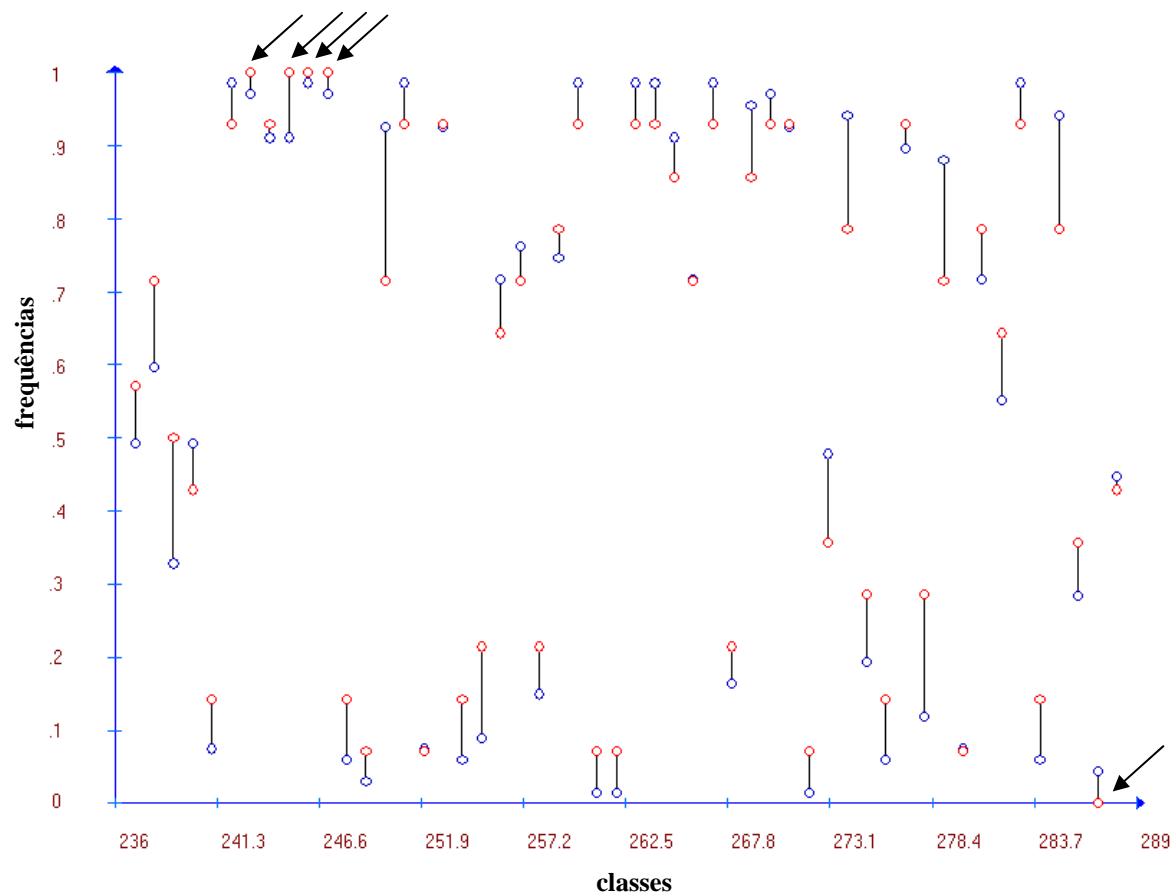


Figura 4. Variação de frequência das 52 classes dos caracteres moleculares. \circ Frequência das classes na coleção inicial, \circ frequência das classes na subcoleção CONV-20. As setas indicam marcas que foram fixadas seja por apresentarem frequência nula ou máxima na subcoleção.

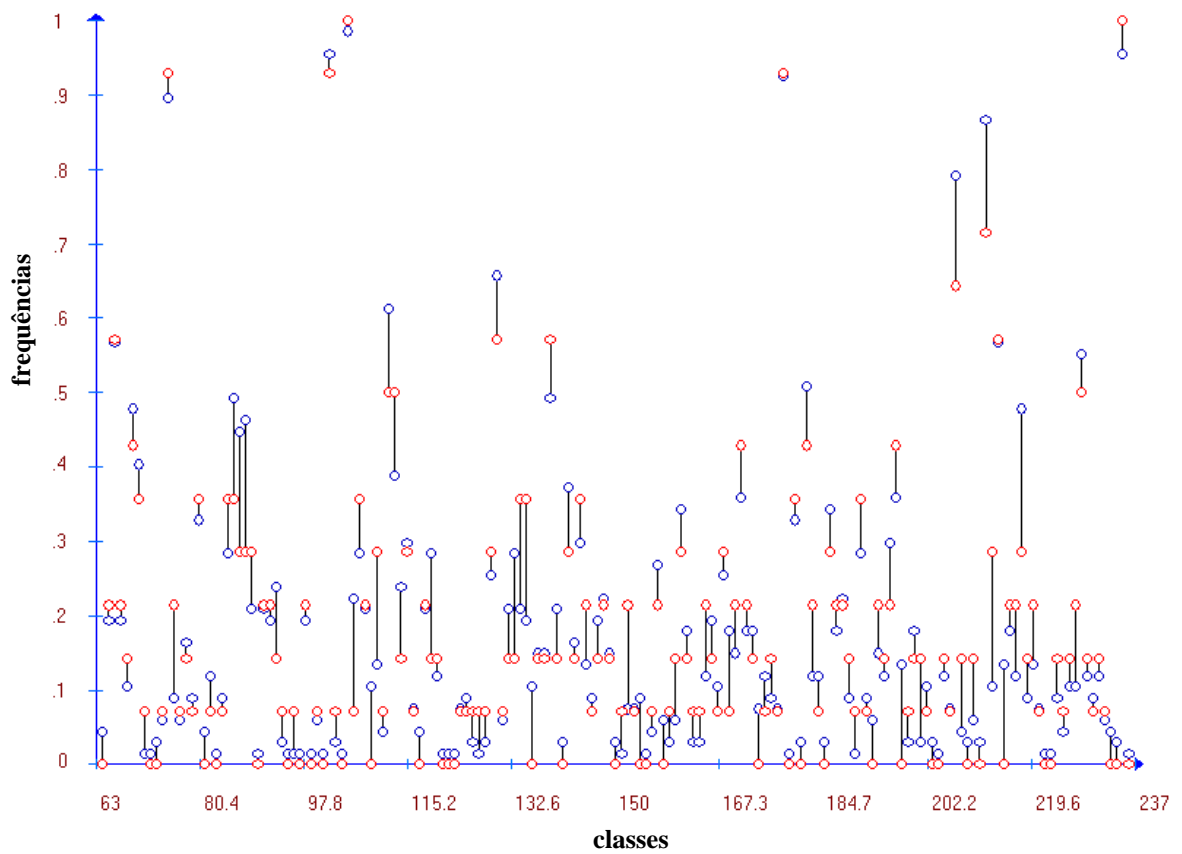


Figura 5. Variação de frequência das 173 classes dos caracteres multicategóricos. ○ Frequência das classes na coleção inicial, ○ frequência das classes na subcoleção MULT-20.

Na análise gráfica das subcoleções MOL-30 e CONV-30 a variabilidade foi mantida para todas as classes de características quantitativas (Figuras 6 e 9), não foram observadas classes com frequência igual a zero.

Já em relação aos caracteres multicategóricos, a subcoleção MOL-30 foi menos eficiente que CONV-30, pois maior número de classes com frequência nula foram observados na subcoleção MOL-30, nesta 29 dos 173 caracteres presentes na coleção inicial não foram representados enquanto em CONV-30 o número de caracteres não amostrados foi 25 (Figuras 7 e 10).

Ao comparar os gráficos de variabilidade das frequências alélicas (classes de caracteres moleculares) para MOL-30 e CONV-30, observou-se a perda de um e três alelos, respectivamente, para estas subcoleções (Figuras 8 e 11). Em se tratando de marcadores moleculares dominantes, a perda de alelos foi contabilizada a partir fixação da presença ou ausência da marca (destacado pelas setas).

De maneira geral, a análise gráfica permitiu concluir que embora o índice coincidência da amplitude e a média das variâncias das características nas subcoleções MOL-30 e CONV-30 indique que ambas são igualmente eficientes, CONV-30 foi ligeiramente superior, pois maior número de classes de caracteres foi mantido no conjunto total em comparação com a subcoleção MOL-30.

Define-se coleção nuclear como um conjunto de subamostras, oriundas de uma coleção de germoplasma, escolhido para representar a máxima variabilidade genética da coleção inicial com o mínimo de redundância (Brown e Spillane, 1999; Vilela-Morales et al., 1997; Cordeiro e Abadie, 2007). Assim, sempre que dados de diferentes naturezas estiverem disponíveis, deve-se priorizar o estabelecimento de coleções nucleares a partir da integração destes dados. Neste contexto, a conversão de dados quantitativos em multicategóricos, utilizando a estratégia DEA-3 foi a mais adequada.

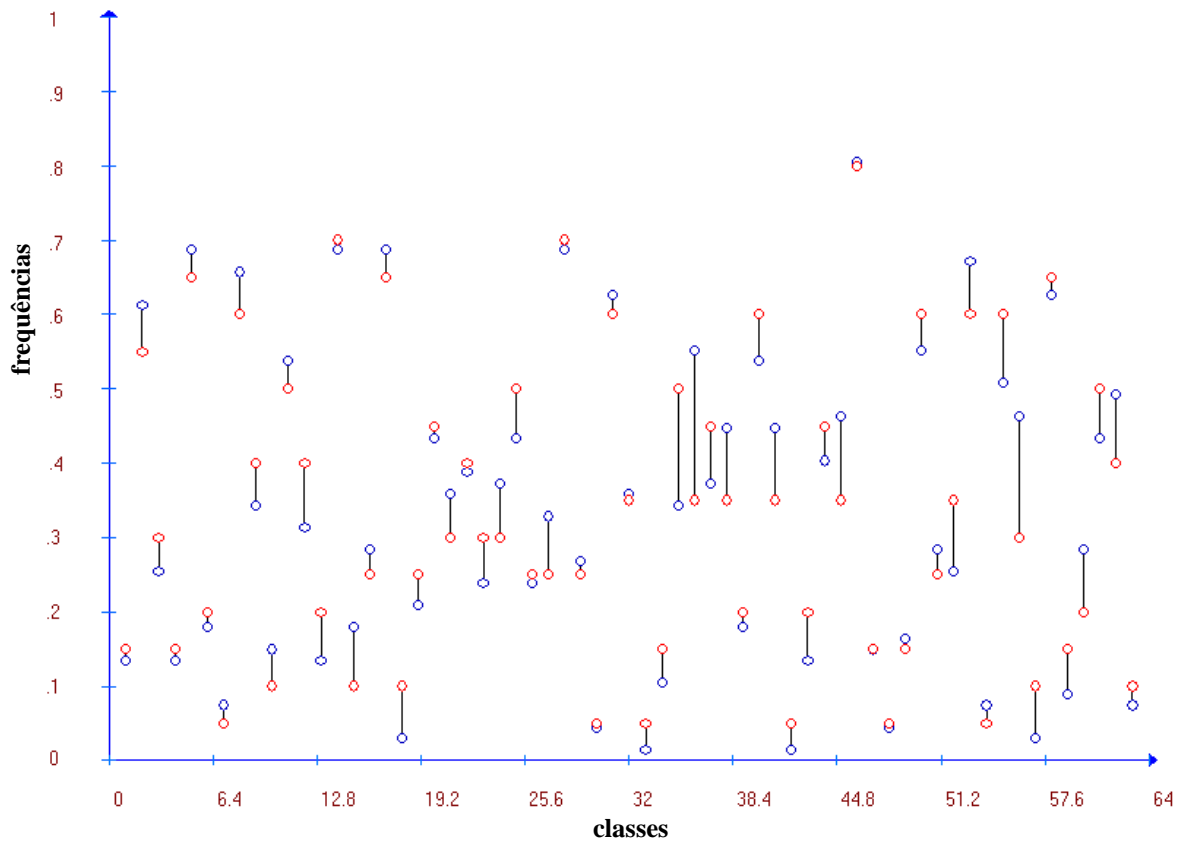


Figura 6. Variação de frequência das 63 classes dos caracteres quantitativos. ○ Frequência das classes na coleção inicial, ○ frequência das classes na subcoleção MOL-30.

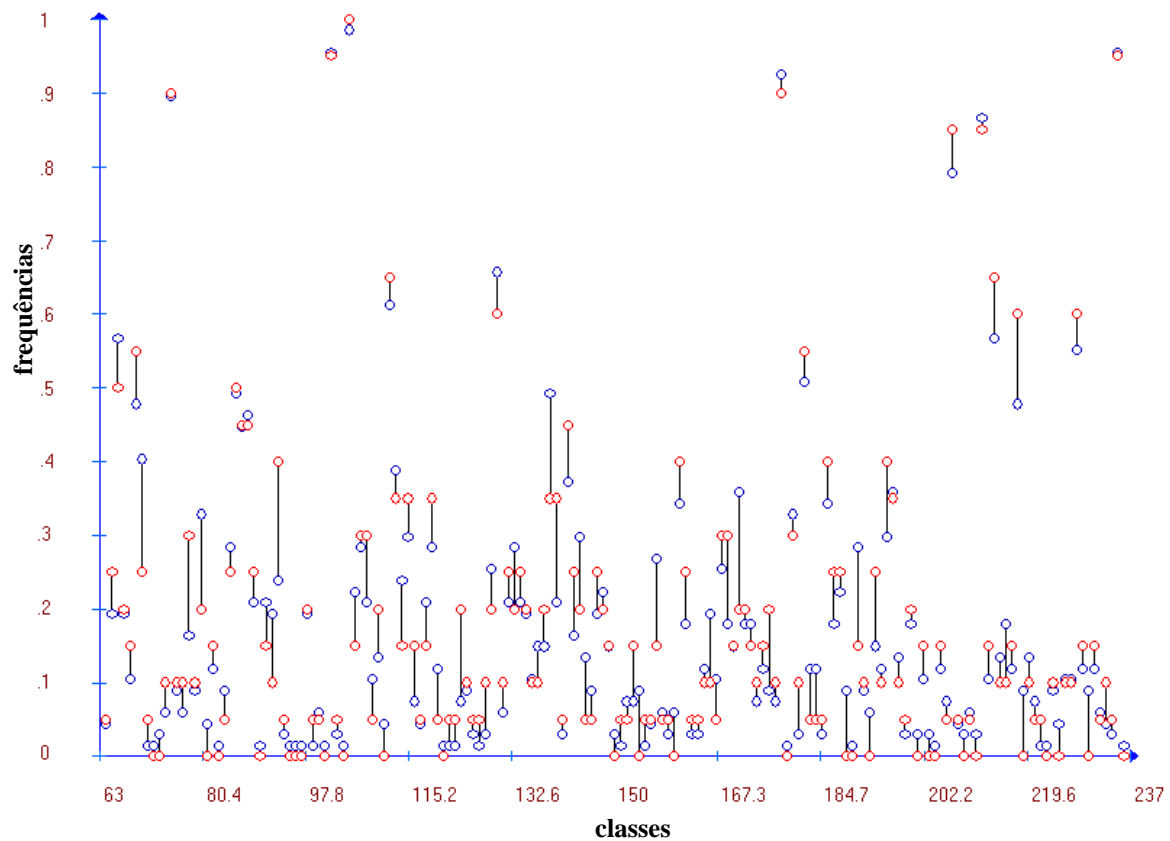


Figura 7. Variação de frequência das 173 classes dos caracteres multicategóricos. ○ Frequência das classes na coleção inicial, ○ frequência das classes na subcoleção MOL-30.

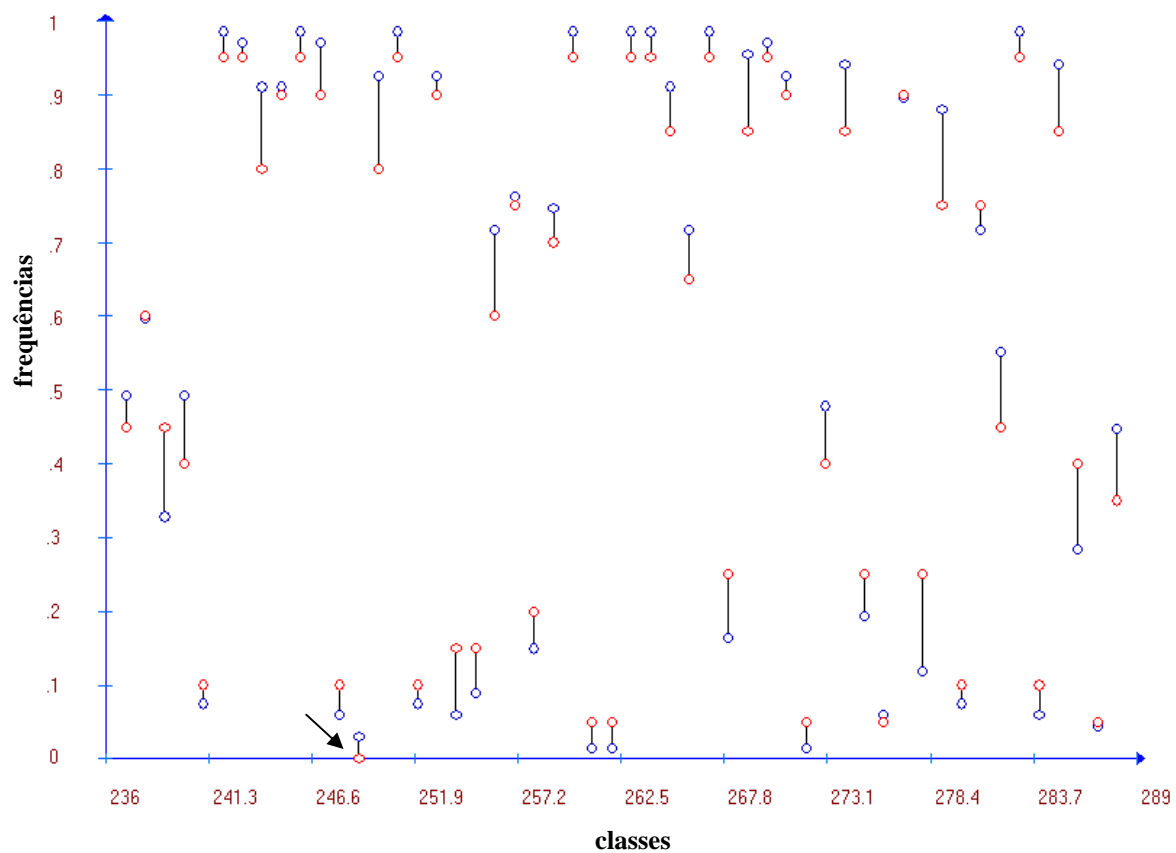


Figura 8. Variação de frequência das 52 classes dos caracteres moleculares. \circ Frequência das classes na coleção inicial, \circ frequência das classes na subcoleção MOL-30. As setas indicam marcas que foram fixadas seja por apresentarem frequência nula ou máxima na subcoleção.

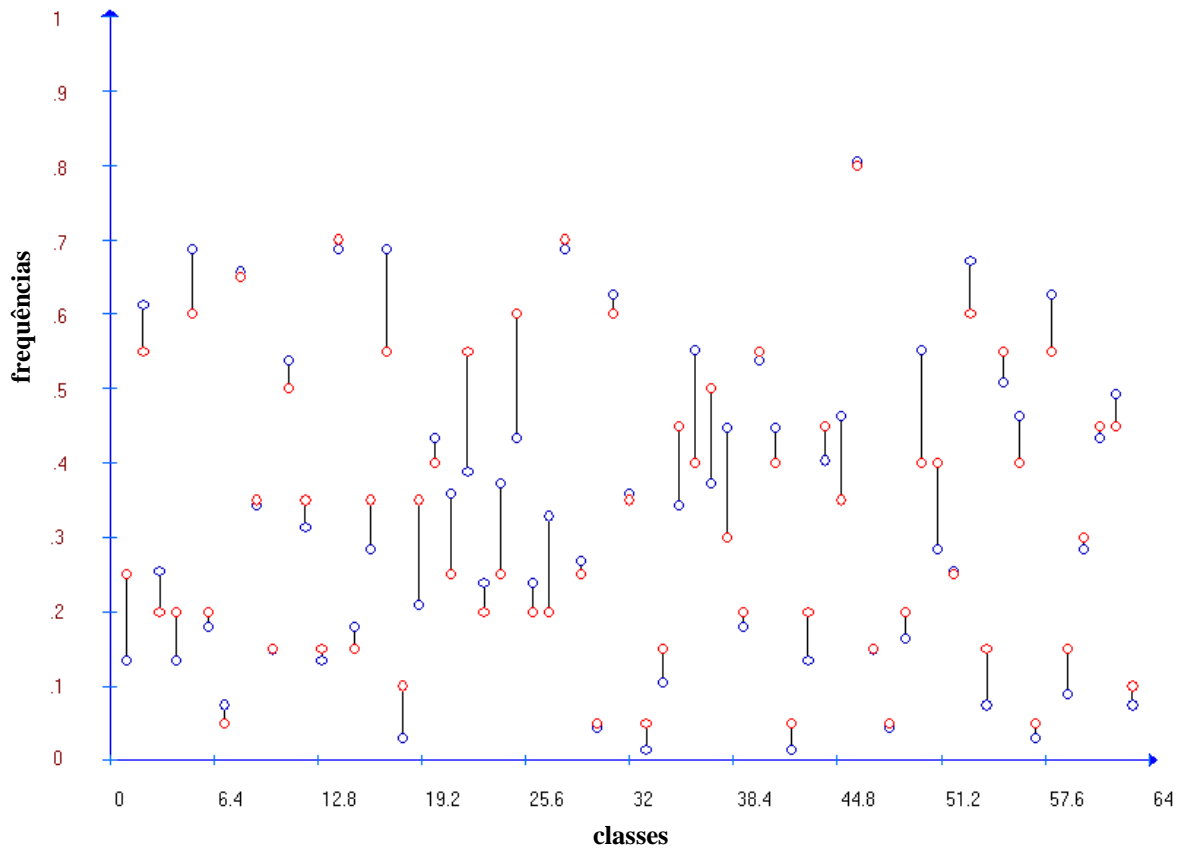


Figura 9. Variação de frequência das 63 classes dos caracteres quantitativos. ○ Frequência das classes na coleção inicial, ○ frequência das classes na subcoleção CONV-30.

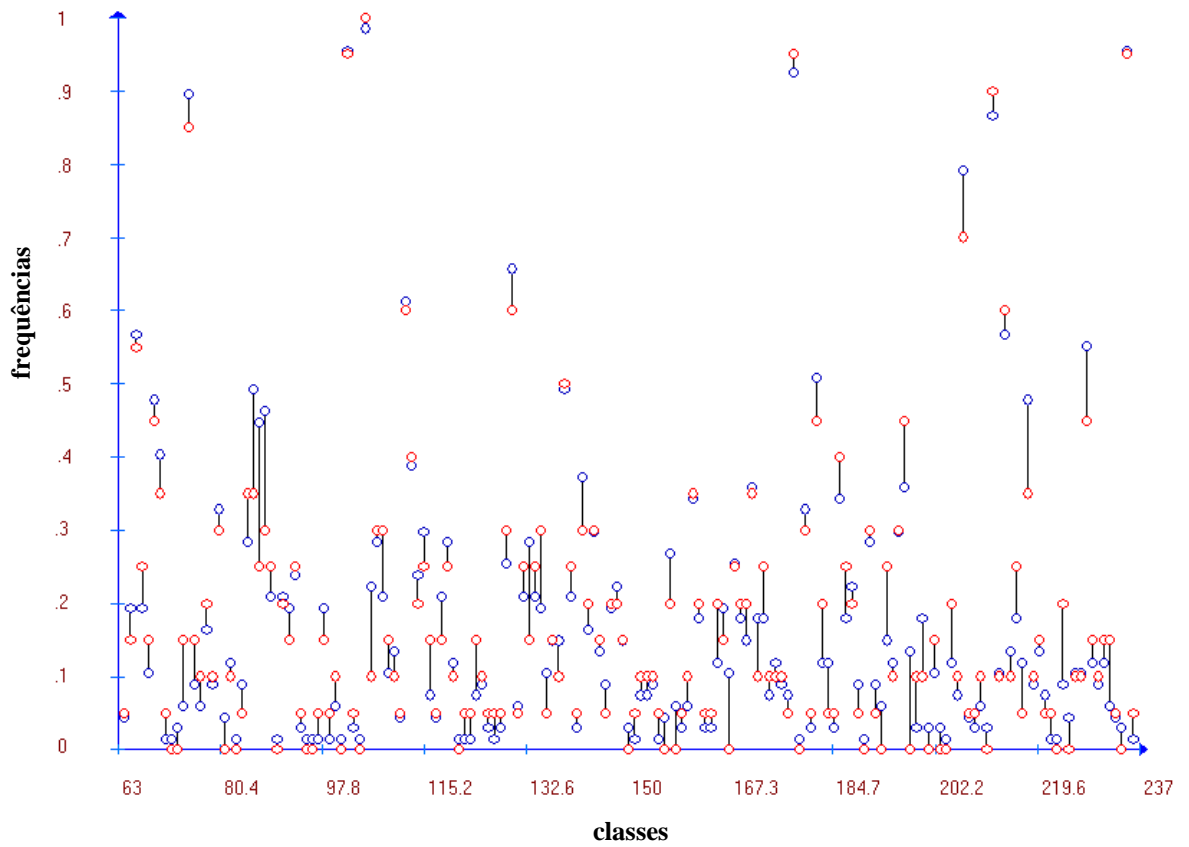


Figura 10. Variação de frequência das 173 classes dos caracteres multicategóricos. ○ Frequência das classes na coleção inicial, ○ frequência das classes na subcoleção CONV-30.

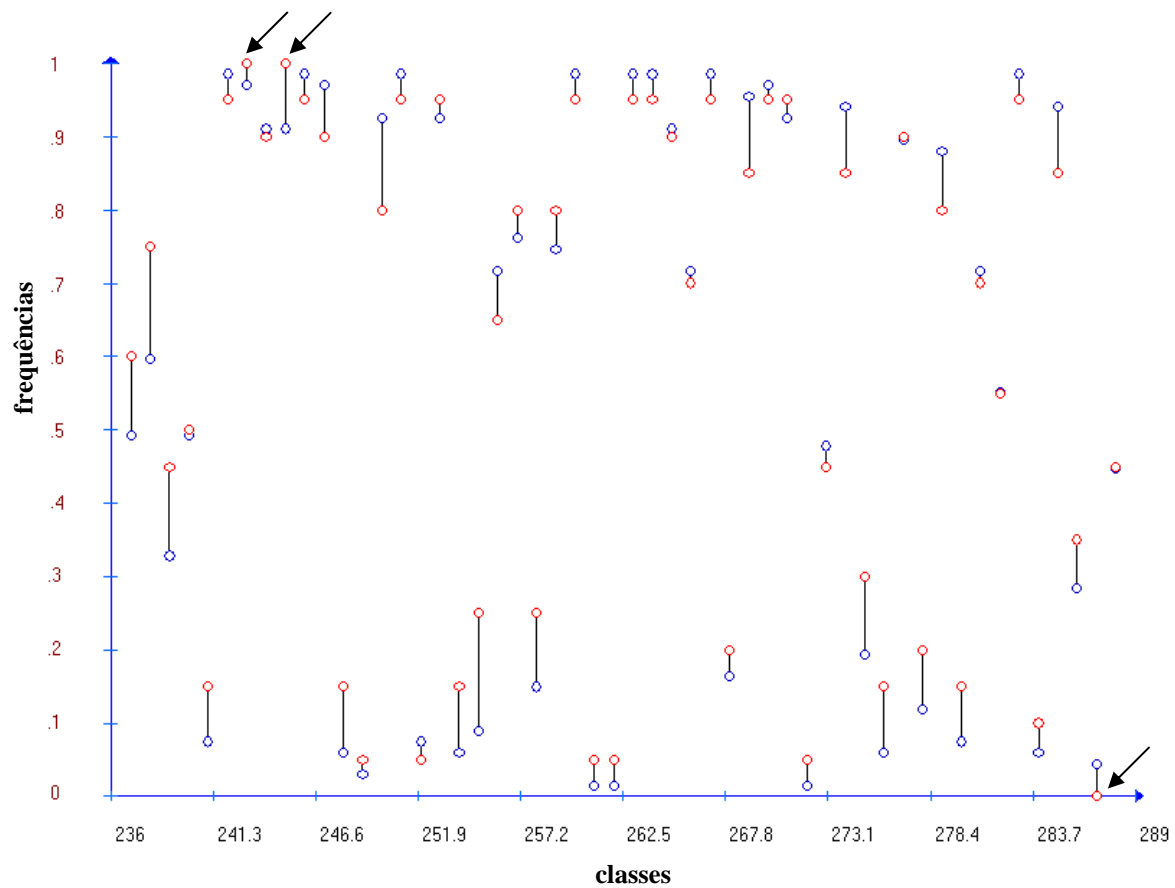


Figura 11. Variação de frequência das 52 classes dos caracteres moleculares. ○ Frequência das classes na coleção inicial, ○ frequência das classes na subcoleção CONV-30. As setas indicam marcas que foram fixadas seja por apresentarem frequência nula ou máxima na subcoleção.

6. CONSIDERAÇÕES FINAIS

Existe grande variabilidade genética entre as subamostras de tomateiro do BGH-UFV estudadas, principalmente em relação aos caracteres quantitativos e multicategóricos avaliados.

O estudo da diversidade genética dos caracteres de diferentes naturezas pelo método de Tocher e comparações das matrizes de dissimilaridade pelo teste Z de Mantel evidenciou uma baixa correlação entre os dados obtidos a partir da análise de caracteres quantitativos, multicategóricos ou moleculares, de forma que as observações em um ou outro estudo não puderam ser extrapoladas, sugerindo-se a necessidade do desenvolvimento de metodologias que possam integrar dados de diferentes naturezas.

Duas estratégias foram utilizadas para a integração dos dados. A primeira foi a conversão dos dados quantitativos e fitopatológicos em multicategóricos, visando a obtenção de uma única matriz de dissimilaridade que contemplasse todas os caracteres independente de suas naturezas. A segunda estratégia utilizada foi a obtenção das matrizes de dissimilaridade individualmente para cada conjunto de caracteres e em seguida a soma algébrica das mesmas. Em relação a conversão de dados, concluiu-se que a melhor estratégia de codificação foi aquela em que o número de classes foi pré-definido, igual a 3, e os limites de cada classe foram estabelecidos a partir da divisão equitativa da amplitude. Quando comparada à estratégia de soma das matrizes, a estratégia de codificação foi considerada melhor uma vez que os grupos formados apresentaram maior

correspondência com sua origem e, ainda, esta estratégia permitiu uma maior discriminação das subamostras, com alguns grupos sendo subdivididos.

A avaliação das diferentes subcoleções quanto à intensidade de amostragem e natureza dos dados revelou coleções bem diferentes, apenas as subamostras 980 e 2234, pertencentes aos estratos Cereja e Maçã, respectivamente, estavam presentes em todas as subcoleções formadas. Apenas as subcoleções baseadas na integração de dados foram validadas para todos os conjuntos de dados, a 20% de intensidade, sendo recomendada a subcoleção CONV-20 por apresentar índices de coincidência da amplitude e variâncias mais adequados. A 30% de intensidade, a subcoleção MOL-30 foi tão eficiente quanto as subcoleções baseadas na integração de dados, quando considerou-se apenas o índice de coincidência da amplitude e a média das variâncias, no entanto a análise gráfica da variabilidade mostrou uma ligeira superioridade da subcoleção CONV-30 em manter a variabilidade, principalmente em relação aos caracteres multicategóricos. Assim, sempre que dados de diferentes naturezas estiverem disponíveis, deve-se priorizar o estabelecimento de coleções nucleares a partir da integração destes dados.

7. REFERÊNCIAS BIBLIOGRÁFICAS

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

ABADIE, T.; CORDEIRO, C. M. T.; FONSECA, J. R., ALVES, R. B. 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**, v. 40, n. 2, p. 129-136, 2005.

ABREU, F. B. LEAL, N. R.; RODRIGUES, R.; AMARAL J. R.; A. T.; SILVA, D. J. H. Divergência genética entre subamostras de feijão-de-vagem de crescimento indeterminado. **Horticultura Brasileira**, v. 22, n. 3, p. 547-552, 2004.

ADALID, A. M.; ROSELLÓ, S.; CONEJO, J. C.; NUEZ, F. Evaluation and selection of *Lycopersicon* accessions for high carotenoid and vitamin c content. **ISHS Acta Horticulture** 789: XV Meeting of EUCARPIA tomato Working Group, 2008.

AGRIANUAL, 2008. **Anuário da agricultura brasileira**. Campo Grande: FNP Consultoria e Comércio, 2008.

AGUILERA, J. G. **Variabilidade molecular e resistência a geminivirus em subamostras de tomateiro do BGH-UFV**. Viçosa, MG: UFV, 2007. 53p. Dissertação (Mestrado em Fitotecnia) – Universidade Federal de Viçosa.

ALLARD, R. W. Predictive methods for germplasm identification. In: Stalker HT, Murphy JP (Ed.) **Plant breeding in the 1990's**. Oxon: Wallingford, 1992. p.119-146.

ALVAREZ, A. E.; VAN de WIEL, C. C. M.; SMULDERS, B. V. Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. **Theoretical and Applied Genetics**, v. 103, p. 1283-1292, 2001.

ALVES, R. M. **Caracterização genética de populações de cupuaçuzeiro, *Theobroma grandiflorum* (Will ex Spreng) Schum., por marcadores microssatélites e descritores botânico-agronômicos**. 2002. 146p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba.

AMARAL JÚNIOR, A. T. do; THIÉBAUT, J. T. de L. **Análise multivariada na avaliação da diversidade em recursos genéticos vegetais**. Campos dos Goytacazes: UENF, 1999. 55 p.

ARAÚJO, M. C. **Uma coleção nuclear de pupunha na amazônia brasileira**. Manaus, AM: UFAM, 2008. 94p. Dissertação (Mestrado em Biotecnologia) – Universidade Federal do Amazonas.

BAR-HEN, A.; CHARCOSSET, A.; BOURGOIN M.; CUIARD J. Relationships between genetic markers and morphological traits in a maize inbred lines collection. **Euphytica**, v. 84, p. 145-154, 1995.

BARONA, G. H.; PARRAS, S. A.; VALLEJO, C. F. A. Evaluation de especies de *Lycopersicon sp.* como fuente de resistencia a *Scrobipalpuloides absoluta* (Meyrick) y si intento de transferencia a *Lycopersicon esculentum* Mill. **Acta Agronomica**, v. 39, n. 1-2, p. 34-45, 1989.

BARRETT, B. A.; KIDWELL, K. K. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. **Crop Science**, v. 38, p. 1261-1271, 1998.

BENTO, C. S.; SUDRE, C. P.; RODRIGUES, R.; RIVA, E. M.; PEREIRA, M. G. Descritores qualitativos e multicategóricos na estimativa da variabilidade fenotípica entre subamostras de pimentas. **Scientia Agraria**, v. 8, n. 2, p. 149-156, 2007.

BERLOO, R. VAN.; ZHU, A.; URSEM, R.; VERBAKEL, H.; GORT, G.; EEUWILK, F. A. VAN. Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. **Theoretical and Applied Genetics**, v. 2008.

BERNARDO, R. Estimation of coefficient of coancestry using molecular markers in maize. **Theoretical and Applied Genetics**, v. 85, p. 1055-1062, 1993.

BREDEMEIJER, G. M. M.; COOKE, R. J.; GANAL, M. W.; PEETERS, R.; ISAAC, P. Construction and testing of a microsatellite database containing more than 500 tomato varieties. **Theoretical and Applied Genetics**, v. 105, p. 1019-1026, 2002.

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**. Cambridge: Cambridge University Press: IPGRI, p. 136-156, 1989.

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

CAMARGO, A. M. M. P.; CAMARGO, F. P.; ALVES, H. S.; CAMARGO-FILHO, W. P. Desenvolvimento do sistema agroindustrial do tomate. **Informações Econômicas**, v. 36, n. 6, p. 53-95, 2006.

CAMARGO-FILHO, W. P.; ALVES, H. S. Ações integradas ao combate à mosca branca do tomate no Brasil. Instituto de Economia Aplicada, 2003. Acesso em dezembro de 2010 <http://www.iea.sp.gov.br/out/Lertexto.php?codTexto=708>

CAMPOS, J. P.; BELFORD, C. C.; GALVÃO, J. D.; FONTES, P. C. R. Efeito da poda da haste e da população de plantas sobre a produção do tomateiro. **Revista Ceres**. Viçosa, v. 34, n. 192, p. 198-208, 1987.

CARELLI, B. P.; GERALD, L. T. S.; GRAZZIOTIN, F. G.; ECHEVERRIGARAY, S. Genetic diversity among Brazilian cultivars and landraces of tomato *Lycopersicon esculentum* Mill. revealed by RAPD markers. **Genetic Resources and Crop Evolution**, v. 53, p. 395-400, 2006.

CARVALHO, J. O. M.; LUZ, J. M. Q.; JULIATTI, F. C.; MELO, L. C.; TEODORO, R. E. F.; LIMA, L. M. L. Desempenho de famílias e híbridos comerciais de tomateiro para processamento industrial com irrigação por gotejamento. **Horticultura Brasileira**, Brasília, v.21, n.3, p. 525-533, 2003.

CHARMET, G.; BALFOURIER, F. The use of geostatistic for sampling a core collection of perennial ryegrass populations. **Genetic Resources and Crop Evolution**, v.42, p.303-309, 1995.

CONSUEGRA, O. P.; MORALES, A. C.; LATERROT, H. E.; ANAÏS, G. B. J. **Mejora genética y manejo del cultivo del tomate para la production em el Caribe**. Habana: Ministério de la Agricultura, 2000. 319p.

CORDEIRO, C. M. T.; ABADIE, T. Coleções nucleares. In: NASS, L. L. (Ed.). **Recursos genéticos vegetais**. Brasília, DF: Embrapa Recursos Genéticos e Biotecnologia, 2007. p. 575-604.

CRUZ, C. D e CARNEIRO, P. C. S. **Modelos Biométricos Aplicados ao Melhoramento Genético**. 2.ed. Viçosa: UFV, 2006. v.2, 585p.

CRUZ, C. D. **Programa Genes**: aplicativo computacional em genética e estatística. Viçosa: UFV, 2006. CD-ROM.

CRUZ, C. D.; REGAZZI, A. J; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa: UFV, 2004. 480 p.

DIWAN, N.; MCINTOSH, M. A. S.; BAUCHAN, G. R. Methods of developing a core collection of annual *Medicago* species. **Theoretical and Applied Genetics**, v.90, p.755-761, 1995.

EGASHIRA, H.; ISHIARA, H.; TAKASHINA, T.; IMANISHI, S. Genetic diversity of the 'peruvianum complex' (*Lycopersicon peruvianum* L.) Mill. and *L. chinense* revealed by RAPD analysis. **Euphytica**, v. 116, p. 23-31, 2000.

EMBRAPA. **Brasil terá o quarto maior banco de germoplasma**. Brasília, 2008. Disponível em: [HTTP://www.cenargen.embrapa.br/cenargen/noticias/anba2904.pdf](http://www.cenargen.embrapa.br/cenargen/noticias/anba2904.pdf).

EMBRAPA. **Situação por hortaliça**. Brasília, 2004. Disponível em: [HTTP://www.cnph.embrapa.br/paginas/hortalicasemnumeros/hotalicasemnumeros.htm](http://www.cnph.embrapa.br/paginas/hortalicasemnumeros/hotalicasemnumeros.htm). Subamostra em 22 de julho de 2008.

FAO. Agriculture production: tomatoes production. Rome 2007. disponível em <http://www.fao.org> acessado em: 15 de dezembro de 2008.

FLORES, M. P. P. **Variabilidade genética de subamostras de tomateiro com base na avaliação de fotossíntese, partição de fotoassimilados e produção**. Viçosa: UFV, 2007, 48p. Dissertação de mestrado em fitotecnia – Universidade Federal de Viçosa, 2007

FONTES, P. C. R.; SILVA, D. J. H. **Produção de tomate de mesa**. 1ª ed. Viçosa: Aprenda fácil, 2002. 193p.

FRALEIGH, B. Global overview of crop genetic resources. In: RUANE, J. & SONNINO, A. (Ed.) **The role of biotechnology in exploring and protecting agricultural genetic resources**. Roma, Itália: Divisão de informação da FAO, v. 10, 2006. pp. 21-31.

FRANCO, J., CROSSA, J.; TABA, S.; EBERHART, S.A. The modified location model for classifying genetic resources: II. Unrestricted variance–covariance matrices. **Crop Science**, v. 42, p. 1727-1736, 2002.

FRANCO, J.; CROSSA, J.; DÍAZ, J.; TABA, S.; VILLASEÑOR, J. e EBERHART, A. A sequential clustering for classifying gene bank accessions. **Crop Science**, 37: 1656-1662, 1997.

FRANCO, J.; CROSSA, J.; RIBAUT, J.M.; BETRAN, J.; WARBURTON, M.L.; KHAIRALLAH, M. A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. **Theoretical and Applied Genetics**, v. 103, p. 944-952, 2001.

FRANCO, J.; CROSSA, J.; VILLASEÑOR, J.; TABA, S. e EBERHART, A. Classifying mexican maize accessions using hierarchical and density search methods. **Crop Science**, 37: 972-980, 1997.

FRANCO, J.; CROSSA, J.; VILLASEÑOR, J.; TABA, S.; EBERHART, S.A. Classifying genetic resources by categorical and continuous variables. **Crop Science**, v. 38, n. 6, p. 1688-1696, 1998.

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 Press, 1984. p. 161-170.

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

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

GOEDERT, C. O. Conservação *ex situ* de recursos genéticos de plantas: caso Embrapa. **Magistra**, v. 18, número especial, p. 15-17, 2006.

GOEDERT, C. O. Histórico e Avanços em Recursos Genéticos no Brasil. In: NASS, L. L. (Ed.) **Recursos Genéticos Vegetais**. Brasília, DF: Embrapa Recursos Genéticos e Biotecnologia. 2007. p. 25-59.

GOMES, C. N. **Caracterização morfo-agronômica e diversidade genética em mandioca *Manihot esculenta* Crantz**. Lavras, MG: UFLA, 2007. 72p. Dissertação (Mestrado em Agronomia/Fitotecnia) – Universidade Federal de Lavras.

GOMES, F. P. Curso de estatística experimental. 4. ed., São Paulo: Livraria Nobel, 1990. 468p.

GRANJA, A. F.; PLATIAU, B.; VARELLA, M. D. Subamostra aos recursos genéticos, transferência de tecnologia e bioprospecção. **Revista Brasileira de Políticas Internacionais**, v. 42, n. 2, p. 81-98, 1999.

HAMRICK, J. L.; GODT, M. J. W. Allozyme diversity in cultivated crops. **Crop Science**, v. 37, p. 26-30, 1997.

HAYS, W. P.; HAYS, R. V. **Foods: the Indians gave us**. New York: Ives Washburn, 1973. 181p.

HOLBROOK, C. C. **Testing and utilization of a core collection for the US germplasm collection of peanut**. In: Johnson RC, Hodgkin T. (Ed.). Core collections for today and tomorrow. Roma: IPGRI, 1999. p. 68-73.

HOLBROOK, C. C.; ANDERSON, W. F. Evaluation of a core collection to identify resistance to late leafspot on peanut. **Crop Science**, v.35, p.1700-1702, 1995.

HOLBROOK, C. C.; ANDERSON, W. F.; PITTMAN, R. N. Selection of a core collection from the US germplasm collection of peanut. **Crop Science**, v.33, p.859-861, 1993.

HOLDEN, J. H. W. The second ten years. In: HOLDEN, J. H. W.; WILIANS, T. J. (Eds.). **Crop genetic resources: conservation and evaluation**. London: G. Allen & Unwin, 1984. p.277-285.

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**, v. 101, p. 264-268, 2000.

IPGRI, **Descriptors for tomato** (*Lycopersicon ssp*). Roma, Itália: International Plant Genetic Resources Institute, 1996. 56p.

JOHNSON, R. C.; HODGKIN, T. **Core collections for today and tomorrow**. Roma: IPGRI, 1999. 81 p.

JUHAZ, A.C.P. **Identificação de fontes de resistência a PepYMV em subamostras de tomateiro (*Lycopersicon spp.*) do Banco de Germoplasma de Hortaliças da UFV, análise da herança e alterações estruturais nos tecidos**. Viçosa: MG: UFV, 2006. 96p. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa.

KARASAWA, M.; RODRIGUES, R.; SUDRÉ, C. P.; SILVA, M. P.; RIVA, E. M.; AMARAL JÚNIOR, A. T. Aplicação de métodos de agrupamento na quantificação da divergência genética entre subamostras de tomateiro. **Horticultura Brasileira**, v.23, n.4, p.1000-1005, 2005.

KOEHLER-SANTOS, P.; DORNELLES, A. L. C.; FREITAS, L. B. de. Characterization of mandarin citrus germplasm from Southern Brazil by morphological and molecular analyses. **Pesquisa Agropecuária Brasileira**, v.38, n.7, p. 747-806, 2003.

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

MARIM, B. G. **Diversidade genética e subcoleção representativa dos subamostras de tomateiro do banco de germoplasma de hortaliças da Universidade Federal de Viçosa**. Viçosa, MG: UFV, 2007. 53p. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa.

MARIM, B.G.; ABREU, F.B.; SILVA, D.J.H.; SAMPAIO JÚNIOR, J.D.; GUIMARÃES, M.A.; LUCA, C.A.C. Dissimilaridade entre subamostras de tomateiro do Banco de Germoplasma de Hortaliças da UFV, utilizando características da fase vegetativa e de produção. **Horticultura Brasileira**, v. 20, n. 2, suplemento 2,

2002. CD-ROM. Trabalho apresentado no 42º Congresso Brasileiro de Olericultura, 2002

MARITA, J. M.; RODRIGUEZ, J. M.; NIENHUIS, J. Development of an algorithm identifying maximally diverse core collections. **Genetic Resources and Crop Evolution**, v. 47, p. 515-526, 2000.

MARSHALL, D. R. Limitations to the use of germplasm collections. In: BROWN, A. H. D.; FRANKEL, O. H.; MARSHALL, D. R.; WILLIAMS, J. T. (Eds.). **The use of Plant Genetic Resources**. Cambridge: Cambridge University Press. 1989. p. 105-120.

MATA, M. C. S.; HURTADO, M. C.; RIPOLLES, S. R.; BALAGUER, L. G.; ISASA, M. E. T.; VIÑALS, F. N. Breeding for flavour of fresh market tomato: sources for increasing acid content. **Acta Physiologiae Plantarum**, v. 22, n. 3, p. 250-253. Apresentado ao XIV Meeting of the EUCARPIA tomato working group, Warsaw: Poland, 2000.

MATSUOKA, K.; CHAVES, G.M. Identificação de raças fisiológicas de *Fusarium oxysporum* f. *lycopersici*, em minas Gerais e seleção de tomateiros resistentes à raça 1 do patógeno. **Experientiae**, v.15, n.10, p.257-289, 1973.

MATTEDI, A. P. **Divergência genética entre subamostras de tomateiro do banco de germoplasma de hortaliças da UFV**. Viçosa, MG: UFV, 2009. 77p. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa.

MESSMER, M. M.; MELCHINGER, A. E.; HERRMANN, R. G.; BOPPENMAIER, J. Relationships among early European maize inbreds: II. Comparison of pedigree and RFLP data. **Crop Science**, v. 33, p. 944–950, 1993.

MILLER J. C; TANKSLEY S. D. RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. **Theoretical and Applied Genetics**, v. 80, p. 437-448, 1990.

MOHAMMADI, A.; PRASANNA, B. M. Analysis of Genetic Diversity in Crop Plants—Salient Statistical Tools and Considerations. **Crop Science**, n. 43, p. 1235-1248, 2003.

N'GORAN, J. A. K.; LAURENT, V.; RISTERUCCI, A M.; LANAUD, C. Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. **Heredity**, v.73, n.6, p.589-597, 1994.

NOLI, E.; CONTI, S.; MACCAFERRI, M.; SANGUINETI, M. C. Molecular characterization of tomato cultivars. **Seed Science Technology**, v. 27, p. 1–10, 1999.

OLIVEIRA, M. do S. P. **Caracterização molecular e morfo-agronômica de germoplasma de açaizeiro**. Lavras, MG. UFLA, 2005. 171p. Tese (Doutorado em Agronomia: Genética e Melhoramento de Plantas) – Universidade Federal de Lavras.

OLIVEIRA, M. F. **Avaliação de cinco estratégias de amostragem para obtenção da coleção nuclear de soja (*Glycine max* (L.) Merrill)**. Piracicaba, SP: ESALQ, 2007. 143p. Tese (Doutorado em Agronomia: Genética e Melhoramento de Plantas) – Escola superior de Agricultura “Luiz de Queiroz”.

PARK, Y. H.; WEST, M. A. L.; ST. CLAIR, D. A. Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum* L.). **Genome**, v. 47, p. 510-518, 2004.

PREGOLATO, W.; PREGOLATO, D. P. **Normas analíticas do Instituto Adolfo Lutz**. 3.ed., São Paulo: Adolfo Lutz, 533p., 1985.

RIBEIRO, N. B.; RODRIGUES, G. S. ; BARRA, V. R.; SILVA, D. J. H.; MIZUBUTI, E. S. G. Desempenho de acessos de tomateiro do Banco de Germoplasma de Hortaliças da UFV, quanto à resistência a *Phytophthora infestans*. **Summa Phytopathologica**, v. 32, suplemento, p. 9-98, 2006.

RICK, C. M. Tomato – *Lycopersicon esculentum* (Solanaceae). In: SMARTT, J.; SIMMONDS, N. W. **Evolution of crop plants**. London: Longman, 1995. p. 449-451.

RICK, C. M. Tomato paste: a concentrated review of genetic highlights from the beginning to the advent of molecular genetics. **Genetics**, v. 128, p. 1-5, 1991.

RODRIGUES, G. S.; RIBEIRO, N. B.; BARRA, V. R.; SILVA, D. J. H., MIZUBUTI, E. S. G. Avaliação de resistência a *Alternaria solani* de subamostras de tomateiro do banco de germoplasma de hortaliças da UFV. In: XXIX Congresso Paulista de Fitopatologia, Botucatu. **Summa Phytopathologica**, v. 32, 2006a.

RODRIGUES, G. S.; RIBEIRO, N. B.; BARRA, V. R.; SILVA, D. J. H.; CARMO, M. G. F.; ROMEIRO, R. S. Caracterização de subamostras de tomateiro do banco de germoplasma de hortaliças da UFV, quanto a resistência a *Ralstonia solanacearum* In: XXIX Congresso Paulista de Fitopatologia, Botucatu. **Summa Phytopathologica**, v. 32, 2006b.

ROHLF, F. J. **Numerical taxonomy and multivariate analysis system**. New York: Exeter Software, 2000. 38 p. (Version 2.1).

SCHULTER, A. R. **Análise isoenzimática, dialéctica e diversidade genética da pimenta silvestre (*Capsicum flexuosum* Sendt.)**. Viçosa: MG. UFV, 1996. 80p. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa.

SCOTT, D. W. On optimal and data-based histograms. **Biometrika**, v. 66, p.605-610, 1979.

SILVA, D. J. H.; MOURA, M. C. C. L.; MARIM, B. G.; ABREU, F. B.; MOREIRA, G. R.; JUASZ, A. C. P.; MATTEDI, A. P.; RIBEIRO, N. B.; AGUILERA, J. G.; FLORES, M. P. Banco de germoplasma de hortaliças – BGH-UFV: caracterização, avaliação e pré-melhoramento. **Magistra**, v. 18, número especial, p. 30-32, 2006.

SILVA, D. J. H.; MOURA, M. C.; CASALI, V. W. D. Recursos genéticos do banco de germoplasma de hortaliças da UFV: Histórico e expedições de coleta. **Horticultura Brasileira**, v. 19, n. 2, p. 108-114, 2001.

SILVA, J. B. C. da; GIORDANO, L. de B.; FURUMOTO, O; BOITEUX, L. da S.; FRANÇA, F. H.; VILLAS BÔAS, G.L.; CASTELO BRANCO, M.; MEDEIROS, M. A.; MAROUELLI, W.; SILVA, W. L. C. E; LOPES, C. A.; ÁVILA, A. C.; NASCIMENTO, W. M.; PEREIRA, W. **Cultivo de tomate para industrialização**. Brasília, DF: Embrapa Hortaliças, 2003. (Sistemas de Produção, 1). Disponível em : <http://sistemasdeprodução.cnptia.embrapa.br/FontesHTML/Tomate/TomateIndustrial/>. Subamostra em: 21 jun. 2007.

SMITH, A. F. **The tomato in América: early, history, culture, and cookery**. Columbia: University of South Carolina Press, 1994. 224p.

SNEDECOR, G. W.; Cochran, W. G. **Statistical Methods**. 7.ed. Ames: Iowa state University, 1980.507p.

SOARES, B. O.; JUHÁSZ, A. C. P. ; SILVA, D. J. H.; OLIVEIRA, J. R.; FUJINAWA M. F. Identificação de fonte de resistência à mancha bacteriana pequena (*Pseudomonas syringae pv. tomato*) em subamostras de tomateiro do Banco de Germoplasma de Hortaliças da UFV. In: Congresso de Brasileiro de Fitopatologia, Maringá. **Fitopatologia Brasileira** (suplemento). v. 32, p. S 156 - S 1556, 2007.

STEVENS, M. A.; RICK, C. M. Genetic and breeding. In: ATHERTON, J. G.; RUDISH, J. (Ed.) **The Tomato Crop**. Cambridge: Chapman and Hall, 1986. p. 35-110.

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

SUINAGA, F. A.; CASALI, V. W. D.; SILVA, D. J. H. da; PICANÇO, M. C. Dissimilaridade genética de fontes de resistência de *Lycopersicon spp.* a *Tuta absoluta* (MEYRICK, 1917) (LEPIDOPTERA: GELECHIDAE). **Revista Brasileira de Agrociência**, v. 9, n. 4, p. 371-376, 2003.

TATINENI, V.; CANTRELI, R. G.; DAVIS, D. D. Genetic diversity in elite cotton germoplasm determined by morfological and RAPDs. **Crop Science**, v.36, n.2, p.186-192, 1996.

TAVARES, C. A. M. Ataque dos vírus. **Cultivar: hortaliças e frutas**, v.2, n.17, p. 26-29, 2003.

VALOIS, A. C. C. Genética aplicada a recursos fitogenéticos. Brasília: Editora UNEB, 1998. 318p.

VAN HINTUM, T. J. L. et al. **Core collections of plant genetic resources**. Roma: IPGRI Technical Bulletin, 2000.

VAN HINTUM, T. J. L. The general methodology for creating a core collection. In: JHONSON, R. C.; HODGKIN, T. (Ed.). **Core collections for today and tomorrow**. Roma: IPGRI, 1999. p. 10-17.

VAN HINTUM, T. J. L.; HAALMAN, D. Pedigree analysis for composing a core collection of modern cultivars, with examples from barley (*Hordeum vulgare*). **Theoretical and Applied Genetics**, v. 88, p. 70-74, 1994.

VASCONCELOS, E. S.; CRUZ, C. D.; BHERING, L. L.; FERREIRA, A. Estratégias de amostragem e estabelecimento de coleções nucleares. **Pesquisa Agropecuária Brasileira**, 42:507-54, 2007.

VILELA-MORALES, E. A.; VALOIS, A. C. C.; NASS, L. L. **Recursos Genéticos Vegetables**. Brasília: EMBRAPA-SPI / EMBRAPA-Cenargen, 1997. 78 p.

VILLAND J., SKROCH ,P. W.; LAI, T.; HANSON, P.; KUO, C. G.; NIEMHUIS, J. Genetic variation among tomato accessions from primary and secondary centers of diversity. **Crop Science**, v. 38, p. 1339–1347, 1998.

WANG, J.; HU, J.; HUANG, X.; XU, S.; ZHANG, S. A strategy on constructing core collections by least distance stepwise sampling. **Theoretical and Applied Genetics**, 115:1-8, 2007.

WETZEL, M. M. V. da S.; MACEDO, J. Organizações nacionais e internacionais de recursos genéticos vegetais. **Magistra**, v. 18, número especial, p. 22-25, 2006.

WILLIAMS C.E.; ST. CLAIR, D.A. Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. **Genome**, v. 36, p. 619-630, 1993.

ZHAO, J.J.; WANG, X.; DENG, B.; PING, L.; WU, J.; SUN, R.; XU, Z.; VROMANS, J.; KOORNNEEF, M.; BONNEMA, G. Genetic relationships within *Brassica rapa* as inferred from AFLP Fingerprints. **Theoretical and Applied Genetics**, v. 110, p. 1301-1314, 2005.