

# Integração de dados em estudos de diversidade genética de tomateiro

Francielle Aline Martins<sup>(1)</sup>, Pedro Crescêncio Souza Carneiro<sup>(2)</sup>, Derly José Henriques da Silva<sup>(3)</sup>, Cosme Damião Cruz<sup>(2)</sup> e José Eustáquio de Souza Carneiro<sup>(3)</sup>

<sup>(1)</sup>Universidade Estadual do Piauí, Avenida Senador Elvídio Nunes Barros, s/nº, CEP 64600-000 Picos, PI. E-mail: franufv@yahoo.com.br

<sup>(2)</sup>Universidade Federal de Viçosa (UFV), Departamento de Biologia Geral, Avenida P. H. Rolfs, s/nº, CEP 36570-000 Viçosa, MG. E-mail: carneiro@ufv.br, cdcruz@ufv.br <sup>(3)</sup>UFV, Departamento de Fitotecnia. E-mail: derly@ufv.br, jesc@ufv.br

Resumo – O objetivo deste trabalho foi integrar dados de caracteres quantitativos, multicategóricos, moleculares e fitopatológicos para a avaliação da diversidade genética de subamostras de tomateiro do Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH-UFV). Foram utilizados dados de 67 subamostras de tomateiro do BGH-UFV, caracterizadas quanto a 19 caracteres quantitativos, 30 multicategóricos, 52 locos ISSR e à reação a três patógenos (*Alternaria solani*, *Pseudomonas syringae* pv. *tomato* e *Tomato yellow spot virus*). Inicialmente, a avaliação da diversidade entre as subamostras foi realizada para cada conjunto de caracteres individualmente, e indicou que a diversidade baseada em qualquer um dos conjuntos de dados não reflete a diversidade dos demais. Para a integração dos dados, codificaram-se os de natureza quantitativa em multicategóricos, por meio de cinco estratégias diferentes. A estratégia de divisão equitativa da amplitude dos dados em três classes foi a mais indicada, com correlação de 0,78 entre as matrizes de dissimilaridade dos dados codificados e originais. A análise de diversidade genética a partir da integração dos dados resultou em grupos com maior correspondência às origens das subamostras de tomateiro avaliadas, o que indica que a integração de dados de diferentes naturezas pode ser realizada com êxito pela conversão dos dados quantitativos em multicategóricos.

Termos para indexação: *Lycopersicon esculentum*, banco de germoplasma, codificação de dados, correlação entre matrizes, método de Tocher.

## Integration of data in studies of genetic diversity of tomato

Abstract – The objective of this work was to integrate data from quantitative, multicategorical, molecular and phytopathological traits in order to evaluate the genetic diversity of tomato subsamples from the germplasm bank of the Universidade Federal de Viçosa (BGH-UFV). Data from 67 tomato subsamples from BGH-UFV were used, characterized according to 19 quantitative traits, 30 multicategorical characters, 52 ISSR loci, and to the reaction to three diseases (*Alternaria solani*, *Pseudomonas syringae* pv. *tomato* and *Tomato yellow spot virus*). Initially, the diversity assessment between the subsamples was performed for each set of characters, indicating that diversity based on any one set of data does not reflect the diversity of another set. For the integration of data, the quantitative ones were coded as multicategorical, by five different strategies. The coding strategy with equitable division of the range of data in three classes was the most appropriate, with correlation of 0.78 between the dissimilarity matrices of encoded and original data. The analysis of genetic diversity by the integration of data resulted in groups with higher correspondence to the origins of the subsamples of tomato, indicating that the integration of data from different sources can be successfully carried out by the conversion of quantitative data in multicategorical.

Index terms: *Lycopersicon esculentum*, germplasm bank, data coding, correlation between matrices, Tocher method.

## Introdução

Os bancos de germoplasma constituem uma forma de conservação ex situ dos recursos genéticos, e são fundamentais para melhorar a produtividade e a sustentabilidade agrícola (Goedert, 2006). Em todo o Brasil, estima-se a existência de 350 bancos de

germoplasma. A disponibilização dos recursos genéticos para o melhoramento de plantas passa necessariamente pela caracterização e pela avaliação agrônômica, fitopatológica, entomológica e molecular da variabilidade genética retida nas coleções de subamostras registradas nos bancos de germoplasma (Silva et al., 2006).

A caracterização e a avaliação dos recursos genéticos da coleção de *Lycopersicon esculentum* Mill. do Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH-UFV) vem sendo realizada de forma sistematizada. As 876 subamostras do BGH-UFV, divididas em lotes, segundo a época de registro, têm sido submetidas à análise mediante o uso dos descritores morfológicos e agrônômicos recomendados pelo International Plant Genetic Resources Institute (1996), com algumas adaptações. Essas subamostras também passam pela caracterização quanto à sensibilidade a pragas e doenças relevantes para o Brasil, e, na última etapa, são avaliadas quanto à variabilidade molecular. A partir desse tipo de caracterização e avaliação, é possível realizar estudos de diversidade genética que incluam a elaboração de uma ou mais coleções nucleares (Silva et al., 2006).

A caracterização e a análise da diversidade genética do germoplasma de uma espécie têm importância fundamental nos programas de melhoramento, pois, a partir dessas avaliações, é possível identificar indivíduos que apresentem genes de interesse agrônômico e inseri-los em programas de cruzamento com cultivares já desenvolvidas.

Vários grupos de dados têm sido utilizados para a análise da diversidade genética em plantas cultivadas. Os principais grupos incluem: dados de pedigree ou passaporte, que frequentemente são incompletos ou limitados (Berloo et al., 2008); dados morfoagronômicos (Marim et al., 2009), que, embora sejam de prática obtenção, sofrem forte influência ambiental e pressão de seleção durante a domesticação e o melhoramento, o que dificulta as interpretações quanto à diversidade (Berloo et al., 2008); e dados de marcadores moleculares, que permitem maior diferenciação dos genótipos (Franco et al., 2001; Zhao et al., 2005).

Estudos de diversidade genética e de elaboração de coleções nucleares têm sido realizados com base em grupos isolados de caracteres, sejam quantitativos, fitopatológicos, multicategóricos ou moleculares. A reunião desses dados em um único estudo é incomum, em parte por existirem poucos bancos de germoplasma que apresentem caracterização detalhada de seu germoplasma. Outro fator que interfere nesse tipo de estudo é a ausência de metodologia de análise eficiente e disponível para a integração dos dados.

O objetivo deste trabalho foi integrar dados de caracteres quantitativos, multicategóricos, moleculares

e fitopatológicos para a avaliação da diversidade genética de subamostras de tomateiro do BGH-UFV.

## Material e Métodos

O estudo foi realizado a partir de um conjunto de dados de 67 subamostras de tomateiro do BGH-UFV: BGH-166, BGH-181, BGH-279, BGH-322, BGH-349, BGH-468, BGH-489, BGH-1254, BGH-4035 e BGH-4206, de procedência não registrada; BGH-773 e BGH-1214, procedentes do Mato Grosso; BGH-850, BGH-4053, BGH-4054 e BGH-4055, procedentes do Espírito Santo; BGH-970, BGH-975, BGH-978, BGH-980, BGH-981, BGH-985, BGH-987, BGH-989, BGH-990, BGH-991, BGH-992, BGH-993, BGH-994, BGH-997, BGH-1490, BGH-1497, BGH-1498, BGH-1499, BGH-1706 e BGH-1708, procedentes de São Paulo; BGH-1019, BGH-1020, BGH-1211, BGH-1532, BGH-4006 e BGH-4309, procedentes de Minas Gerais; BGH-1282, procedente do Paraná; BGH-1485, procedente do Rio Grande do Sul; BGH-1538, procedente de Goiás; e as subamostras de procedência norte-americana, BGH-1985, BGH-1987, BGH-1988, BGH-1989, BGH-1990, BGH-1991, BGH-1992, BGH-1993, BGH-2119, BGH-2202, BGH-2203, BGH-2205, BGH-2208, BGH-2211, BGH-2213, BGH-2214, BGH-2216, BGH-2219, BGH-2223, BGH-2229, BGH-2234 e BGH-3472.

Tradicionalmente chamada de acesso, a subamostra refere-se à porção do material biológico devidamente acompanhada de informações biológicas, químicas ou documentais, que permitam a identificação taxonômica e de procedência do material registrado em banco de germoplasma. Conforme orientação técnica nº 2, de 30 de outubro de 2003, do Conselho de Gestão do Patrimônio Genético (Brasil, 2003), o termo subamostra deve ser utilizado para designar germoplasmas que foram registrados em banco de germoplasma, uma vez que o termo acesso apresenta conotação legal, com sentido distinto ao que vem sendo utilizado.

Todas as 67 subamostras foram previamente caracterizadas quanto a 19 caracteres quantitativos, 30 caracteres multicategóricos, 53 locos ISSR – determinados por Aguilera et al. (2011) –, e quanto à reação a *Alternaria solani*, *Pseudomonas syringae* pv. *tomato* e ao begomovírus *Tomato yellow spot virus* (ToYSV).

As características quantitativas foram avaliadas nas diferentes fases do desenvolvimento do tomateiro.

Na fase de plântula, avaliou-se o diâmetro do hipocótilo (DH) e o comprimento do cotilédone (CC). Na fase vegetativa, foram avaliados a espessura do pecíolo principal (EPP) e os comprimentos da folha (CFo) e do entrenó (CE). As características dos frutos avaliadas foram: comprimento do fruto (CFr), largura do fruto (LFr) e do eixo central (LEC), espessura do endocarpo (EE), número de lóculos (NL), teor de sólidos solúveis totais (SST), acidez total titulável (ATT) e qualidade organoléptica (QO). Avaliaram-se, ainda, características agrônômicas, como: peso (PFB) e número de frutos bons (NFB), peso (PTF) e número total de frutos (NTF), peso médio de frutos (PMF) e índice de precocidade (IP).

Os dados de reação das subamostras a *A. solani* e *P. syringae* também são de natureza quantitativa, pois são obtidos, respectivamente, a partir da medida da área foliar lesionada pelo fungo e da contagem do número total de pústulas bacterianas em cada planta. A reação das subamostras ao ToYSV foi caracterizada em cinco classes, como: altamente resistente (AR), resistente (R), moderadamente resistente (MR), suscetível (S) e altamente suscetível (AS).

Os dados dessa avaliação foram analisados com os dados multicategóricos: cor do hipocótilo (CH); tipo de crescimento da planta (TCP); densidade de pilosidade na haste (DPH) e de folhagem (DF); atitude (AtF) e tipo de folha (TF); tipo de florescimento da corola (TFC); coloração externa do fruto imaturo (CEFI); presença e frequência de ombros verdes no fruto (POVF); formato (FPF), homogeneidade (HTF) e tamanho do fruto (TFr); coloração (CEFM) e intensidade de coloração (ICEFM) do fruto maduro; formato secundário (FSF) e formato do ombro do fruto (FOF); medida da área de cortiça ao redor da cicatriz do pedicelo (MACRCP); facilidade de destacar o epicarpo (FDE); cor do epicarpo (CorE); cor (CM) e intensidade de cor do mesocarpo (IM); formato da sessão transversal do fruto (FSTF) e da cicatriz estilar (FCE); forma da extremidade distal do fruto (FEDF); condição da cicatriz estilar (CCE); cor do eixo central (CEC); rachaduras radial (RR) e concêntrica (RC).

Inicialmente, o estudo da diversidade genética foi realizado para cada conjunto de dados. Em relação aos dados quantitativos, a matriz de dissimilaridade foi obtida a partir da distância euclidiana média padronizada. Para os dados multicategóricos, assim como para os dados moleculares, a matriz de dissimilaridade foi

obtida por meio do complemento aritmético do índice de coincidência simples. A existência de correlação entre as matrizes de dissimilaridade foi verificada pelo teste Z de Mantel e pelo teste t, a 5% de probabilidade. A partir de cada uma das matrizes de dissimilaridade, procedeu-se à análise de diversidade genética pelo método de agrupamento de Tocher, conforme Cruz & Carneiro (2006).

Para a integração dos dados de diferentes naturezas, codificaram-se os de natureza quantitativa em multicategóricos a partir de cinco estratégias, que variaram quanto ao número de classes e ao critério de estabelecimento dos limites entre classes (K). Na estratégia de divisão equitativa da amplitude (DEA), as médias das características quantitativas foram agrupadas em 2, 3, 4, 6, 8 ou 10 classes. Os limites de cada classe foram estabelecidos a partir da DEA. Na estratégia de percentual equitativo dos dados (PED), o número de classes foi pré-definido em 2, 3, 4, 6, 8 ou 10. Cada uma das classes foi estabelecida de forma que todas apresentassem a mesma quantidade de dados. Na estratégia com número fixo de classes (DN), as médias dos dados quantitativos foram subdivididas em seis classes, tendo-se levado em consideração a média e o desvio-padrão (DP), de acordo com o seguinte critério: classe 1, mínimo, média menos dois DP; classe 2, média menos dois DP e média menos um DP; classe 3, média menos um DP e média; classe 4, média e média mais um DP; classe 5, média mais um DP e média mais dois DP; classe 6, máximo, média mais dois DP. Na estratégia de definição de classes, segundo o teste de Scott-Knott (SK), o número de classes (K) foi estabelecido de acordo com o número de grupos formados para cada característica pelo teste, a 5% de probabilidade. Na estratégia de Scott (S), o número de classes foi estabelecido a partir da equação  $K = (A \times n^{1/3}) / (3,49 \times s)$ , em que: A é a amplitude total, s é o desvio-padrão de cada característica e n é o número total de amostras, conforme Scott (1979).

Após as conversões, novas matrizes de dissimilaridade foram obtidas para cada estratégia, tendo-se utilizado como medida de dissimilaridade o complemento aritmético do índice de coincidência simples. Foram estimados os coeficientes de correlação entre essas matrizes e a matriz de dissimilaridade obtida dos dados quantitativos originais, a 5% de probabilidade, pelo teste Z de Mantel e pelo teste t, para a determinação da forma mais adequada de conversão dos dados

quantitativos. Em seguida, realizou-se a análise da diversidade genética pelo método de agrupamento de Tocher, a partir de uma única matriz de dissimilaridade que contemplou as informações quantitativas, multicategóricas e moleculares, simultaneamente.

As análises de diversidade, as correlações entre matrizes e a codificação de dados quantitativos foram realizadas com auxílio do programa Genes (Cruz, 2006).

## Resultados e Discussão

O agrupamento das 67 subamostras de tomateiro do BGH-UFV pelo método de Tocher (Tabela 1) resultou em 14 grupos, tanto para dados quantitativos, com uso da distância euclidiana média padronizada como medida de dissimilaridade, quanto para os multicategóricos, com uso da matriz de distância, obtida a partir do complemento do índice de coincidência simples. Entretanto, a distribuição das subamostras entre os grupos foi diferenciada. No agrupamento pelo método de Tocher, com dados quantitativos, observou-se maior concentração de subamostras nos grupos 1 e 2. Já no agrupamento com uso do conjunto de dados multicategóricos, a distribuição das subamostras entre os 14 grupos foi mais equitativa.

No agrupamento pelo método de Tocher, a partir de informações moleculares, observou-se que a diversidade molecular é mais restrita, embora haja diversidade fenotípica. Todas as 67 subamostras

foram agrupadas em apenas dois grupos. O grupo 2 foi formado somente pela subamostra de número 980, enquanto as demais foram alocadas no grupo 1.

Segundo Gomes (2007), a formação de número diferente de grupos e a distribuição diferenciada das subamostras nos grupos, mesmo quando o número de grupos formados é o mesmo, indicam discordância entre os procedimentos de agrupamento realizados. Assim, as inferências em relação à diversidade entre as 67 subamostras de tomateiro do BGH-UFV, com base em cada conjunto de caracteres (quantitativo, multicategórico ou molecular), não podem ser extrapoladas às de outro conjunto.

Para cada par de matrizes de dissimilaridade, obtidas com base nos grupos de caracteres, foi estimado o coeficiente de correlação (Tabela 2), pelo teste t e pelo teste Z de Mantel. Foram observados valores de correlação relativamente baixos. Segundo Rohlf (2000), valores de correlação inferiores a 0,7 têm sido considerados inconsistentes, o que indica que a diversidade genética obtida com base em um conjunto de caracteres não explica a diversidade com base em outro conjunto. A significância de valores de correlação de baixa magnitude pode ser atribuída ao elevado número de graus de liberdade (2.209), pelo teste t, e ao elevado número de observações (2.211) comparadas nas matrizes de dissimilaridade, pelo teste Z de Mantel.

A decisão de qual grupo de variáveis utilizar para determinação da divergência genética e de qual medida

**Tabela 1.** Agrupamento de 67 subamostras de tomateiro do Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa, pelo método de Tocher, com base na dissimilaridade dos caracteres quantitativos e multicategóricos.

| Grupo | Dados quantitativos                                                                                                                         | Dados multicategóricos                                                                 |
|-------|---------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| 1     | 975, 991, 987, 990, 992, 970, 1532, 985, 279, 993, 994, 166, 322, 850, 1706, 1538, 4206, 4055, 1497, 1499, 1498, 1708, 1490, 349, 1020, 181 | 993, 994, 970, 992, 975, 1020, 985, 279, 990, 1254, 991, 850, 322, 2111, 987, 166, 980 |
| 2     | 2214, 2219, 3472, 2223, 2216, 2213, 2205, 1987, 2208, 2211, 2202, 997, 1485, 2229, 2219, 1985, 981, 978, 1019, 1992, 1214                   | 1497, 1990, 1708, 1706, 1498, 1706, 1499, 1490, 1538, 1532, 2119, 1992, 1485           |
| 3     | 4006, 4035, 4053, 4309                                                                                                                      | 2216, 2229, 2211, 2219, 2223, 2205, 2208, 2202                                         |
| 4     | 1988, 1991, 1282, 1989                                                                                                                      | 4006, 4053, 4035, 4206, 4055, 4054, 4309                                               |
| 5     | 489, 773, 1254                                                                                                                              | 981, 1019, 997, 349, 978, 989, 468, 1214                                               |
| 6     | 468                                                                                                                                         | 2213, 2234                                                                             |
| 7     | 980                                                                                                                                         | 489, 773                                                                               |
| 8     | 1211                                                                                                                                        | 1985, 1991, 1988, 1282                                                                 |
| 9     | 1993                                                                                                                                        | 2203, 3472                                                                             |
| 10    | 2203                                                                                                                                        | 2214                                                                                   |
| 11    | 4054                                                                                                                                        | 1987                                                                                   |
| 12    | 1990                                                                                                                                        | 1993                                                                                   |
| 13    | 989                                                                                                                                         | 181                                                                                    |
| 14    | 2234                                                                                                                                        | 1989                                                                                   |

de dissimilaridade adotar, depende do objetivo do estudo. A utilização de caracteres multicategóricos é prática, econômica e demanda menor tempo, em comparação a caracteres quantitativos e moleculares (Marim et al., 2009). Em muitas culturas, caracteres quantitativos, mesmo sendo de difícil mensuração, têm sido preferidos em estudos de diversidade, por apresentar importância comercial (Gomes, 2007). Já a utilização de caracteres moleculares tem sido vantajosa, principalmente, em virtude de não serem influenciados pelo ambiente (Berloo et al., 2008).

O uso de distâncias de dissimilaridade de diferentes conjuntos de dados têm produzido 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 coeficiente de correlação igual a 0,09. Este autor atribuiu a inexistência de correlação entre as duas medidas de dissimilaridade à diferença do controle genético nos diferentes tipos de caracteres analisados. Oliveira (2005), ao comparar dissimilaridades genéticas e fenotípicas de açaizeiro, verificou valores de correlação inferiores a 0,05. Koehler-Santos et al. (2003) também não obtiveram concordância no padrão de variação quando compararam as matrizes geradas com os dados morfológicos e os marcadores microsatélites, ao avaliar germoplasma de *Citrus*. No entanto, Alves (2002) observou coeficiente de correlação igual a 0,72, ao avaliar a associação entre 21 locos microsatélites 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, e, geralmente, é baixa quando a amostra é pequena.

**Tabela 2.** Coeficientes de correlação (r) entre matrizes de dissimilaridade, a partir de dados de caracteres quantitativos, multicategóricos e moleculares.

| Matrizes comparadas                | r         |
|------------------------------------|-----------|
| Quantitativos vs. multicategóricos | 0,46808** |
| Quantitativos vs. moleculares      | 0,20946** |
| Multicategóricos vs. moleculares   | 0,21379** |

\*\*Significativo a 1% de probabilidade, pelo teste t e também pelo teste de Mantel, baseado em 1.000 simulações, nas comparações de matrizes de dados quantitativos vs. multicategóricos e de dados multicategóricos vs. moleculares.

Portanto, a diversidade observada a partir de um conjunto de dados não pode ser extrapolada aos demais, em virtude dos baixos valores de correlação entre as matrizes de dissimilaridade, para cada conjunto de dados (Tabela 2). Os três conjuntos de caracteres são importantes para a organização da diversidade do germoplasma, principalmente nos estudos de diversidade genética que tenham como objetivo a escolha de subamostras para compor uma coleção nuclear, para conservação da variabilidade. Neste caso, é necessário identificar uma forma de integrar informações de diferentes naturezas, em uma única análise.

No total, foram obtidas 15 matrizes de dissimilaridade a partir da conversão dos dados quantitativos em multicategóricos. Os valores de correlação entre a matriz de dados originais e as matrizes obtidas pela codificação dos dados foram consistentes apenas para as estratégias DEA, com número de classes igual a 2, 3 e 4 (Tabela 3).

A divisão em classes, conforme a estratégia PED, não foi eficiente, pois não foram observados valores de correlação consistentes entre as matrizes obtidas a partir do uso dessa estratégia e a matriz de dados originais.

**Tabela 3.** Coeficiente de correlação (r) entre a matriz de dados quantitativos originais e as matrizes obtidas por diferentes estratégias de recodificação dos dados.

| Estratégia <sup>(1)</sup> | Nº de classes                         | r      |
|---------------------------|---------------------------------------|--------|
| DEA                       | 2                                     | 0,73** |
|                           | 3                                     | 0,78** |
|                           | 4                                     | 0,71** |
|                           | 6                                     | 0,66** |
|                           | 8                                     | 0,59** |
|                           | 10                                    | 0,56** |
| PED                       | 2                                     | 0,63** |
|                           | 3                                     | 0,61** |
|                           | 4                                     | 0,55** |
|                           | 6                                     | 0,51** |
|                           | 8                                     | 0,46** |
|                           | 10                                    | 0,42** |
| DN                        | 6                                     | 0,66** |
| SK                        | Variou de acordo com a característica | 0,66** |
| S                         | Variou de acordo com a característica | 0,68** |

<sup>(1)</sup>DEA, divisão equitativa da amplitude; PED, percentual equitativo dos dados; DN, número de classes fixo, pressupondo distribuição normal dos dados; SK, número de classes estabelecido pelo teste de Scott-Knott; S, número de classes estabelecido conforme Scott (1979). \*\*Significativo a 1% de probabilidade, pelo teste t e pelo teste de Mantel, baseado em 1.000 simulações.

Observou-se número ideal de classes para a categorização dos dados quantitativos. Quando as classes são estabelecidas a partir da estratégia DEA, para esse conjunto de dados do BGH-UFV, o número de classes ideal foi 3, cujo valor de correlação entre a matriz dos dados convertidos e originais foi de 0,78. Contudo, não foi possível determinar número ideal de classes para a estratégia PED, tendo-se observado que quanto maior o número de classes utilizadas na codificação, menor foi a correlação.

Os métodos mais utilizados para a distribuição de valores quantitativos em classes são a regra de Sturges (1926) e a do quadrado. Para o conjunto de dados avaliado no presente trabalho, o número de classes, de acordo com essas regras, foi de 7 e 8, respectivamente. No entanto, a codificação em seis ou mais classes não é consistente, mesmo quando a amplitude de cada classe é estabelecida de forma equitativa.

O número de classes utilizadas na codificação variou para cada característica, com uso das estratégias SK e S. Na estratégia SK, as subamostras foram classificadas para as variáveis DH, EPP e ATT, em duas classes; para CC, CE, PFB e reação a *Alternaria*, em três classes; para QO, em quatro classes; para CFo, LFr, EE, NL, NFB, NTF e SST, em cinco classes; para PTF e IP, em seis classes; para LEC e reação a *Pseudomonas*, em sete classes; e para CFr e PMF, em oito classes. Já na estratégia S, as subamostras foram classificadas quanto à reação a *Alternaria*, em três classes; a EE e LEC, em quatro classes; a EPP, LFr, NL, PFB, PMF e PTF, em cinco classes; a DH, CC, CE, NFB, NTF, IP, SST, ATT, QO e quanto à reação a *Pseudomonas*, em seis classes; e a CFo e CFr, em sete classes. A correlação

entre as matrizes de dados originais e as matrizes de dados codificados pelas estratégias SK e S também não foram consistentes.

Após definida a melhor subestratégia de codificação dos dados (DEA-3), procedeu-se ao estudo de diversidade genética com a integração dos dados de naturezas diferentes. Assim, a análise de diversidade genética foi realizada ao se considerar 104 caracteres (21 quantitativos codificados, 31 multicategóricos e 52 moleculares). O agrupamento das 67 subamostras pelo método de Tocher foi baseado em uma única matriz de dissimilaridade obtida a partir do complemento aritmético do índice de coincidência simples (Tabela 4).

Foram alocadas, no grupo 1, cerca de 46% das subamostras, todas de procedência nacional. Destacase que o segundo grupo foi formado apenas por subamostras norte-americanas. Isso se repetiu na formação do grupo 4, com exceção da subamostra BGH-1485, que foi alocada nesse grupo, cuja procedência é o Rio Grande do Sul. As subamostras BGH-181, BGH-1993 e BGH-980 foram alocadas nos grupos 8, 9 e 10, respectivamente.

O estudo da diversidade genética do tomateiro, a partir da integração dos dados de diferentes naturezas, pelo método de agrupamento de Tocher, resultou em grupos com maior correspondência em relação às procedências das subamostras, o que indica a consistência da análise de diversidade genética quando se aumentou o número de caracteres considerados. Uma vez integrados, os dados de diferentes naturezas podem ser utilizados na definição das subamostras que deverão compor a coleção nuclear, para conservação da variabilidade.

**Tabela 4.** Agrupamento de 67 subamostras de tomateiro do Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa, pelo método de Tocher, a partir da integração de dados por meio da estratégia divisão equitativa da amplitude em três classes.

| Grupo | Subamostras                                                                                                                                                                 |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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                                                                                                                                                                         |

## Conclusões

1. Embora haja diversidade fenotípica entre as subamostras de tomateiro do BGH-UFV, a diversidade molecular observada é restrita.

2. As inferências baseadas em qualquer dos conjuntos de dados, de natureza quantitativa, multicategórica ou molecular, não podem ser extrapoladas para outro conjunto, na avaliação da diversidade genética de subamostras de tomateiro.

3. Na codificação de dados quantitativos em multicategóricos, destaca-se a estratégia de divisão equitativa da amplitude em três classes.

4. A integração de dados de diferentes naturezas para estudo de diversidade genética pode ser realizada com êxito pela conversão dos dados quantitativos em multicategóricos, com uso do complemento aritmético do índice de coincidência simples como medida de dissimilaridade.

## Referências

- AGUILERA, J.G.; PESSONI, L.A.; RODRIGUES, G.B.; ELSAYED, A.Y.; D.J.H. da; BARROS, E.G. de. Genetic variability by ISSR markers in tomato (*Solanum lycopersicon* Mill.). **Revista Brasileira de Ciências Agrárias**, v.6, p.243-252, 2011.
- ALVES, R.M. **Caracterização genética de populações de cupuaçuzeiro, *Theobroma grandiflorum* (Willd. ex. Spreng.) Schum., por marcadores microssatélites e descritores botânico-agronômicos**. 2002. 146p. Tese (Doutorado) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.
- BERLOO, R. van; ZHU, A.; URSEM, R.; VERBAKEL, H.; GORT, G.; EEUWILK, F.A.V. Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. **Theoretical and Applied Genetics**, v.117, p.89-101, 2008.
- BRASIL. Ministério do Meio Ambiente. Conselho de Gestão do Patrimônio Genético. Orientação técnica nº 1, de 24 de setembro de 2003. Esclarece os conceitos de acesso e de remessa de amostras de componentes de patrimônio genético. **Diário Oficial [da] República Federativa do Brasil**, 24 out. 2003. Seção 1, p.79.
- CRUZ, C.D. **Programa Genes**: aplicativo computacional em genética e estatística. Viçosa: UFV, 2006. 442p.
- CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 2.ed. Viçosa: UFV, 2006. 585p.
- 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.
- GOEDERT, C.O. Conservação *ex situ* de recursos genéticos de plantas: caso Embrapa. **Magistra**, v.18, p.15-17, 2006.
- GOMES, C.N. **Caracterização morfo-agronômica e diversidade genética em mandioca *Manihot esculenta* Crantz**. 2007. 72p. Dissertação (Mestrado) – Universidade Federal de Lavras, Lavras.
- INTERNATIONAL PLANT GENETIC RESOURCES INSTITUTE. **Descriptors for tomato (*Lycopersicon* ssp.)**. Roma: IPGRI, 1996. 56p.
- 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, p.797-806, 2003.
- MARIM, B.G.; SILVA, D.J.H. da; CARNEIRO, P.C.S.; MIRANDA, G.V.; MATTEDI, A.P.; CALIMAN, F.R.B. Variabilidade genética e importância relativa de caracteres em acessos de germoplasma de tomateiro. **Pesquisa Agropecuária Brasileira**, v.44, p.1283-1290, 2009.
- 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, p.589-597, 1994.
- OLIVEIRA, M. do S.P. **Caracterização molecular e morfo-agronômica de germoplasma de açaizeiro**. 2005. 171p. Tese (Doutorado) – Universidade Federal de Lavras, Lavras.
- ROHLF, F.J. **NTSYS-pc: numerical taxonomy and multivariate analysis system**. New York: Exeter Software, 2000. 38p.
- 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.; YUASZ, A.C.P.; MATTEDI, A.P.; RIBEIRO, N.B.; AGUILERA GONZÁLEZ, J.; FLORES, M.P. Banco de germoplasma de hortaliças – BGH-UFV: caracterização, avaliação e pré-melhoramento. **Magistra**, v.18, p.30-32, 2006.
- STURGES, H.A. The choice of a class interval. **Journal of the American Statistical Association**, v.21, p.65-66, 1926.
- 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.

Recebido em 12 de janeiro de 2011 e aprovado em 23 de outubro de 2011