

# Tamanho de coleção original, métodos de agrupamento e amostragem para obtenção de coleção nuclear de germoplasma

Edmar Soares de Vasconcelos<sup>(1)</sup>, Cosme Damião Cruz<sup>(2)</sup>, Derly José Henriques da Silva<sup>(3)</sup>  
e Leonardo Lopes Bhering<sup>(2)</sup>

<sup>(1)</sup>Universidade Estadual do Oeste do Paraná, Centro de Ciências Agrárias, Rua Pernambuco, nº-1.777, Caixa Postal 91, CEP 85960-000 Marechal Cândido Rondon, PR. E-mail: [vasconceloses@pop.com.br](mailto:vasconceloses@pop.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: [cdcruz@ufv.br](mailto:cdcruz@ufv.br), [leonardo.bhering@ufv.br](mailto:leonardo.bhering@ufv.br) <sup>(3)</sup>UFV, Departamento de Fitotecnia. E-mail: [derly@ufv.br](mailto:derly@ufv.br)

**Resumo** – O objetivo deste trabalho foi verificar o efeito do tamanho da coleção original de germoplasma sobre eficácia de diferentes métodos de agrupamento e de amostragem, utilizados na obtenção de coleções nucleares, e comparar esses métodos no estabelecimento de coleções nucleares. Foram simulados sete tamanhos de coleções originais e utilizadas sete estratégias de amostragem para estabelecimento de coleções nucleares, com uso de caracteres morfo-agronômicos. Oito estratégias de obtenção de coleção nuclear foram empregadas, seis com técnicas de agrupamento (Tocher e Tocher sequencial) e duas sem agrupamento (aleatória e Tocher invertido). Nas estratégias de agrupamento, foram empregadas as amostragens constante, logarítmica e proporcional. Determinaram-se a variância explicada, o valor do coeficiente de determinação, o coeficiente de coincidência e o coeficiente de coincidência do desvio-padrão entre a coleção nuclear e a coleção original. As estratégias de amostragem constante e logarítmica geram coleções nucleares representativas da coleção original. O agrupamento de Tocher sequencial é mais eficaz na representação da coleção original pela coleção nuclear do que o agrupamento de Tocher. Entre as estratégias avaliadas, a amostragem pelo método de Tocher com critério de aglomeração inverso foi a mais eficiente na geração de coleções nucleares representativas das coleções originais.

**Termos para indexação:** agrupamento de acessos, estratégias de amostragem, método de Tocher inverso, recursos genéticos.

## Size of original collection, grouping and sampling methods for the obtention of germplasm core collection

**Abstract** – The objective of this study was to verify the effect of original germplasm collection size on the effectiveness of different clustering methods and sampling, used to obtain core collections, and to compare these methods in the establishment of core collections. Simulations were made for seven sizes of original collections, and seven sampling strategies were used for the establishment of core collections using morpho-agronomic traits. Eight strategies for obtaining core collections were used, six of them by grouping techniques (Tocher and sequential Tocher), and two without grouping (random and Tocher with inverted criteria grouping). For grouping strategies, constant, logarithmic and proportional sampling were used. The explained variance, the value of the determination coefficient, the coincidence coefficient, and the coincidence standard deviation coefficient between the core and the original collections were determined. The constant and logarithmic sampling strategies generate core collections which are representative of the original collections. The sequential Tocher grouping is more efficient in representing the original collection by the core collection than the Tocher's. Among the evaluated strategies the sampling by Tocher's inversed grouping criterion is the most efficient to generate representative core collections of the original collections.

**Index terms:** accession grouping, sampling strategies, Tocher's inverse method, genetic resources.

### Introdução

O grande tamanho das coleções de germoplasma é frequentemente obstáculo à sua utilização, conservação e manejo. Coleções como a de arroz do International

Rice Research Institute (IRRI), com mais de 100 mil acessos, e a de sorgo do International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), com mais de 36.700 acessos, além do custo elevado para sua manutenção, são de difícil manuseio. A utilização

dos acessos de enormes coleções na maioria dos casos, também fica comprometida.

As coleções nucleares (core collections) são constituídas por um grupo limitado de acessos, derivados de uma coleção de germoplasma, escolhidos para representar a variabilidade genética da coleção inteira (Brown, 1989). Por se tratar de um número reduzido de acessos, seu manuseio tem menor custo e os acessos podem ser melhor caracterizados.

No estabelecimento de coleções nucleares, podem ser usados dados de caracteres moleculares e morfológicos (Ghislain et al., 1999; Gouesnard et al., 2001). Para esse estabelecimento, as estratégias utilizadas, de maneira geral, fazem uso de um método de agrupamento e um critério de amostragem dentro de cada grupo (Grenier et al., 2001; Chandra et al., 2002; Holbrook & Dong, 2005).

O método de agrupamento de Tocher tem sido amplamente empregado no estudo da divergência genética entre acessos. Oliveira et al. (2004), Silva et al. (2005) e Souza et al. (2005) utilizaram esse método, para avaliar a divergência genética entre genótipos de melancia, alface e cana-de-açúcar, respectivamente. Contudo, alterações no processo de agrupamento de Tocher têm sido sugeridas.

Vasconcelos et al. (2007b) sugerem uma alteração no método de Tocher, para incorporar características do método de agrupamento sequencial, para que os genótipos de um grupo não interfiram no agrupamento dos demais. Isso facilita o processo de amostragem dentro de cada grupo, já que dificilmente se terá um grupo formado por apenas um genótipo.

Diferentes critérios podem ser empregados para a amostragem e elaboração da coleção nuclear. Grenier et al. (2001) utilizaram as seguintes estratégias: número constante de acessos selecionados de cada grupo, número de acessos selecionados proporcional ao tamanho do grupo formado, e número de acessos igual à razão do logaritmo do número de acessos de cada grupo dividido pela soma dos logaritmos de cada grupo.

Outros estudos também têm sido realizados para avaliar as diferentes estratégias de amostragem (Chandra et al., 2002; Franco et al., 2005). Chandra et al. (2002) avaliaram cinco estratégias de amostragem: a aleatória, a constante, a proporcional, a logarítmica e a raiz quadrada. Para seus dados simulados de isoenzimas

em batata, as estratégias proporcional e aleatória foram as mais indicadas.

Há, ainda, um método que não faz uso de agrupamento seguido de amostragem – o método de agrupamento de Tocher com critério de aglomeração inverso (Vasconcelos et al., 2007a) – que consiste em formar um único grupo, pela seleção dos acessos que permitem obtenção da maior divergência dentro da coleção original e que passariam a constituir a coleção nuclear.

O objetivo deste trabalho foi verificar o efeito do tamanho da coleção original de germoplasma sobre a eficácia de diferentes métodos de agrupamento e de amostragem, utilizados na obtenção de coleções nucleares, e comparar esses métodos no estabelecimento de coleções nucleares.

## Material e Métodos

Foram simuladas sete coleções de germoplasma com tamanhos de 100, 250, 500, 750, 1.000, 1.500 e 2.000 acessos. Os acessos dos bancos de germoplasma foram descritos em relação a 21 características morfoagronômicas (Tabela 1), tendo-se tomado por base os dados do estudo de Huaman et al. (1999), em que foram avaliados acessos de batata-doce, *Ipomoea*

**Tabela 1.** Médias e variâncias das 21 características das coleções originais simuladas.

Característica	Média	Variância
1. Entrelaçamento	0,6600	0,0049
2. Hábito de crescimento	6,8200	0,0025
3. Diâmetro do internódio da rama	3,8000	0,0016
4. Comprimento dos internódios da rama	2,8300	0,0036
5. Cor predominante da rama	4,6100	0,0064
6. Cor da rama secundária	2,0000	0,0081
7. Esboço geral da folha	4,5600	0,0025
8. Tipo de lóbulo da folha	3,7000	0,0100
9. Número de lóbulos da folha	3,7200	0,0036
10. Forma do lóbulo da folha central	2,9700	0,0049
11. Tamanho da folha madura	5,2500	0,0009
12. Pigmentação das nervuras abaxiais da folha	4,9400	0,0064
13. Cor da folha madura	3,0300	0,0036
14. Cor da folha imatura	3,4100	0,0036
15. Pigmentação do pecíolo	5,1000	0,0121
16. Comprimento do pecíolo	4,2900	0,0025
17. Cor predominante da película externa da raiz	13,1600	0,0900
18. Cor secundária da película externa da raiz	1,5700	0,0121
19. Cor predominante da raiz	3,7600	0,0036
20. Cor secundária da raiz	2,6500	0,0144
21. Distribuição da cor secundária da raiz	1,4800	0,0081

*batatas* (L.). Os parâmetros de média e variância foram fixados e simulados independentemente do tamanho das coleções.

O tamanho das coleções nucleares foi definido com base no trabalho de Vasconcelos et al. (2007a), pelo método de agrupamento de Tocher com aglomeração inversa. As coleções nucleares, obtidas para as coleções de 100 e 250 acessos, foram formadas por amostragem de 10%, ou seja, 10 e 25 entradas respectivamente. As coleções nucleares das demais coleções (500, 750, 1.000, 1.500 e 2.000) foram formadas por 5% dos acessos (25, 38, 50, 75 e 100 entradas, respectivamente).

Foram calculadas as distâncias euclidianas entre os acessos de germoplasma das coleções, baseadas nos dados de caracterização simulados. A partir dessas distâncias, foram realizados os agrupamentos dos acessos pelos métodos apresentados a seguir.

Para comparar os métodos de agrupamento, e dividir a coleção original em grupos de acessos a serem amostrados, foram empregadas duas metodologias distintas: o método de agrupamento de Tocher (Cruz & Carneiro, 2006) e o método de Tocher sequencial (Vasconcelos et al., 2007b).

No método de otimização de Tocher, realiza-se a partição do conjunto de acessos em subgrupos não vazios e mutuamente exclusivos, por meio da maximização ou minimização de uma medida preestabelecida, em que se adota o valor máximo da medida de dissimilaridade – encontrado no conjunto das menores distâncias que envolvem cada progenitor (designado por  $\alpha$ ) – como o limite de acréscimo na média da distância intragrupo, ou seja, para a formação ou inclusão de acesso a um grupo.

No método de Tocher sequencial, a obtenção de  $\alpha$  e a formação do grupo 1 é realizada da mesma forma que no de Tocher; porém, para a formação do grupo 2, deve ser obtido um novo valor de  $\alpha$ , assim, volta-se ao início da divisão em grupos (antes da formação do grupo 1). Isso irá propiciar um novo valor do critério de agrupamento após a formação de cada novo grupo, e não um critério único, conforme o método de Tocher original.

Avaliaram-se, ainda, diferentes estratégias de amostragem. Nos grupos obtidos por cada critério de agrupamento, foram aplicados diferentes esquemas de amostragem: aleatória, dentro de cada grupo; proporcional ao número de acessos de cada grupo;

e com base na relação do logaritmo do tamanho do grupo, pela soma do logaritmo de todos os grupos.

O método de formação de uma coleção nuclear, descrito por Vasconcelos et al. (2007b), também foi avaliado. Nesse método, a seleção dos acessos para compor as coleções nucleares usa a técnica de agrupamento de Tocher (Cruz & Carneiro, 2006), porém, com uma modificação que consiste em agrupar acessos com maior dissimilaridade, ou seja, método de Tocher com critério de aglomeração inverso. Por meio dessa alteração, os acessos selecionados formam um único grupo que corresponde à coleção nuclear.

Para obter uma menção da eficiência das estratégias de agrupamento e amostragem para a formação da coleção nuclear, a amostragem dos acessos foi realizada de forma aleatória sem a formação de grupos ou estratos, que seria o convencional. Dessa forma, para cada coleção, foram obtidas oito coleções nucleares. Dois métodos de agrupamento, com três estratégias de amostragem (seis coleções nucleares), uma coleção nuclear pela estratégia do método de Tocher inverso, e outra pela seleção aleatória das entradas.

Após a obtenção das diferentes coleções nucleares, avaliou-se a eficácia delas em representar as coleções de germoplasma, com um teste de homogeneidade entre as variâncias (teste F) e um teste entre as médias (teste t) de cada característica da coleção original e das respectivas características nas coleções nucleares.

Foi estabelecido, também, o coeficiente de similaridade de média e de variância entre cada coleção e sua respectiva coleção original. De acordo com Hu et al. (2000), o coeficiente de similaridade média pode ser obtido por:

$$CR (\%) = \left( \frac{1}{m} \sum_{j=1}^m \frac{R_c}{R_I} \right) 100,$$

em que: CR(%) é o coeficiente de similaridade entre a coleção original e sua coleção nuclear; m é o número de características avaliadas;  $R_c$  é a diferença entre o valor máximo e o mínimo de cada variável, observado na coleção nuclear;  $R_I$  é a diferença entre o valor máximo e o mínimo de cada variável, observado na coleção original.

O coeficiente de similaridade da variância pode ser obtido por:

$$CRV (\%) = \left( \frac{1}{m} \sum_{j=1}^m \frac{V_c}{V_I} \right) 100,$$

em que:  $CRV(\%)$  é o coeficiente de similaridade da variância entre a coleção original e sua respectiva coleção nuclear;  $m$  é o número de características avaliadas;  $V_c$  é a variância de cada variável, observada na coleção nuclear;  $V_i$  é a variância de cada variável, observada na coleção original.

Dois outros critérios, utilizados para verificar a eficácia de representatividade das coleções nucleares, foram a percentagem da variância da coleção original, explicada pela coleção nuclear ( $S^2$  explicada), e o coeficiente de determinação ( $R^2$ ) determinado pela razão entre a soma de quadrados dos desvios, entre as entradas da coleção nuclear e a soma de quadrados total dos desvios dos acessos da coleção original (Cruz, 2006).

As simulações das coleções originais, a obtenção das distâncias euclidianas, a formação dos grupos, a amostragem dos acessos pelo método de Tocher com critério de aglomeração inverso, e a comparação de médias e variâncias entre as coleções foram efetuadas com auxílio do aplicativo computacional Genes (Cruz, 2006).

## Resultados e Discussão

As médias das 21 características das diferentes coleções nucleares foram, de maneira geral, idênticas às das coleções originais pelo teste t, a 5% de probabilidade (Tabela 2). Contudo, em alguns casos e para no máximo três características, a coleção nuclear não apresentou média semelhante à da coleção original. Isto foi verificado para a coleção nuclear obtida mediante o método de agrupamento de Tocher

sequencial, com amostragem constante na população original de 1.000 acessos.

As médias das variáveis avaliadas na coleção nuclear foram, em geral, semelhantes às das respectivas coleções originais, já que entre 18 e 21 características, das 21 avaliadas nas coleções nucleares e originais, foram estatisticamente semelhantes. Com exceção do método de Tocher com critério invertido de aglomeração, todas as demais estratégias de agrupamento empregadas tiveram médias significativamente diferentes entre as coleções originais e as respectivas coleções nucleares.

Com exceção do método de Tocher invertido, as estratégias empregadas no presente trabalho, para a obtenção de uma coleção nuclear, geraram coleções nucleares com variância das 21 características avaliadas semelhantes às variâncias das características da coleção original.

O método de Tocher com critério de aglomeração invertido possibilitou obter coleções nucleares com, no mínimo, o dobro de variância média das coleções originais, o que é explicável.

A padronização dos dados, da variância de uma característica  $x$  [ $V(x)$ ] será dada por  $V(x) = (\sum_{i=1}^n x_i^2)/n-1$ . A obtenção de uma coleção nuclear reduz o valor de  $n$ . Se  $\sum_{i=1}^n x_i^2$  diminui na mesma proporção de  $n$ , a variância da coleção nuclear será próxima à da coleção original, como verificado nos métodos de agrupamento de Tocher e de Tocher sequencial e, também, para a estratégia de seleção aleatória. Contudo, quando a coleção nuclear é obtida pelo método de Tocher com critério de aglomeração inverso, por utilizar a

**Tabela 2.** Número de médias (MD) e número de variâncias (VD), das 21 variáveis de cada coleção nuclear que diferiram das médias da coleção de germoplasma original. Variância média (VA) de cada coleção original e de suas respectivas coleções nucleares formadas por diferentes estratégias.

Estratégia de seleção	Pop100 ( $S^2=0,127$ ) <sup>(5)</sup>			Pop250 ( $S^2=0,112$ ) <sup>(5)</sup>			Pop500 ( $S^2=0,100$ ) <sup>(5)</sup>			Pop750 ( $S^2=0,106$ ) <sup>(5)</sup>			Pop1.000 ( $S^2=0,097$ ) <sup>(5)</sup>			Pop1.500 ( $S^2=0,101$ ) <sup>(5)</sup>			Pop2.000 ( $S^2=0,096$ ) <sup>(5)</sup>		
	MD	VD	VA	MD	VD	VA	MD	VD	VA	MD	VD	VA	MD	VD	VA	MD	VD	VA	MD	VD	VA
Aleatório	1	2	0,070	1	0	0,072	1	1	0,079	1	4	0,099	0	2	0,085	1	1	0,088	0	1	0,081
Tocher aleatório	0	1	0,113	2	2	0,084	0	1	0,091	0	2	0,105	0	2	0,085	0	2	0,098	0	3	0,103
Tocher logaritmo	0	1	0,115	0	1	0,076	2	1	0,088	0	2	0,116	1	0	0,091	2	0	0,096	0	2	0,110
Tocher prop. <sup>(1)</sup>	0	2	0,086	0	2	0,080	0	2	0,072	1	1	0,097	0	2	0,078	0	0	0,116	1	3	0,085
Tocher seq. a. <sup>(2)</sup>	0	2	0,130	0	6	0,210	0	1	0,086	0	3	0,110	3	2	0,132	0	2	0,101	0	12	0,140
Tocher seq. log. <sup>(3)</sup>	0	2	0,261	0	3	0,168	1	2	0,137	1	1	0,099	2	5	0,099	1	3	0,090	0	6	0,122
Tocher seq. p. <sup>(4)</sup>	1	1	0,223	2	1	0,097	1	1	0,111	0	1	0,095	0	2	0,094	0	3	0,108	0	0	0,090
Tocher invertido	0	4	0,576	0	11	0,351	0	15	0,301	0	15	0,276	0	20	0,251	0	21	0,216	0	21	0,230

<sup>(1)</sup>Agrupamento de Tocher, com amostragem proporcional ao tamanho do grupo. <sup>(2)</sup>Agrupamento de Tocher sequencial, com amostragem constante. <sup>(3)</sup>Agrupamento de Tocher sequencial, com amostragem logarítmica. <sup>(4)</sup>Agrupamento de Tocher sequencial, com amostragem proporcional ao tamanho do grupo. <sup>(5)</sup>Variância da referida população simulada.

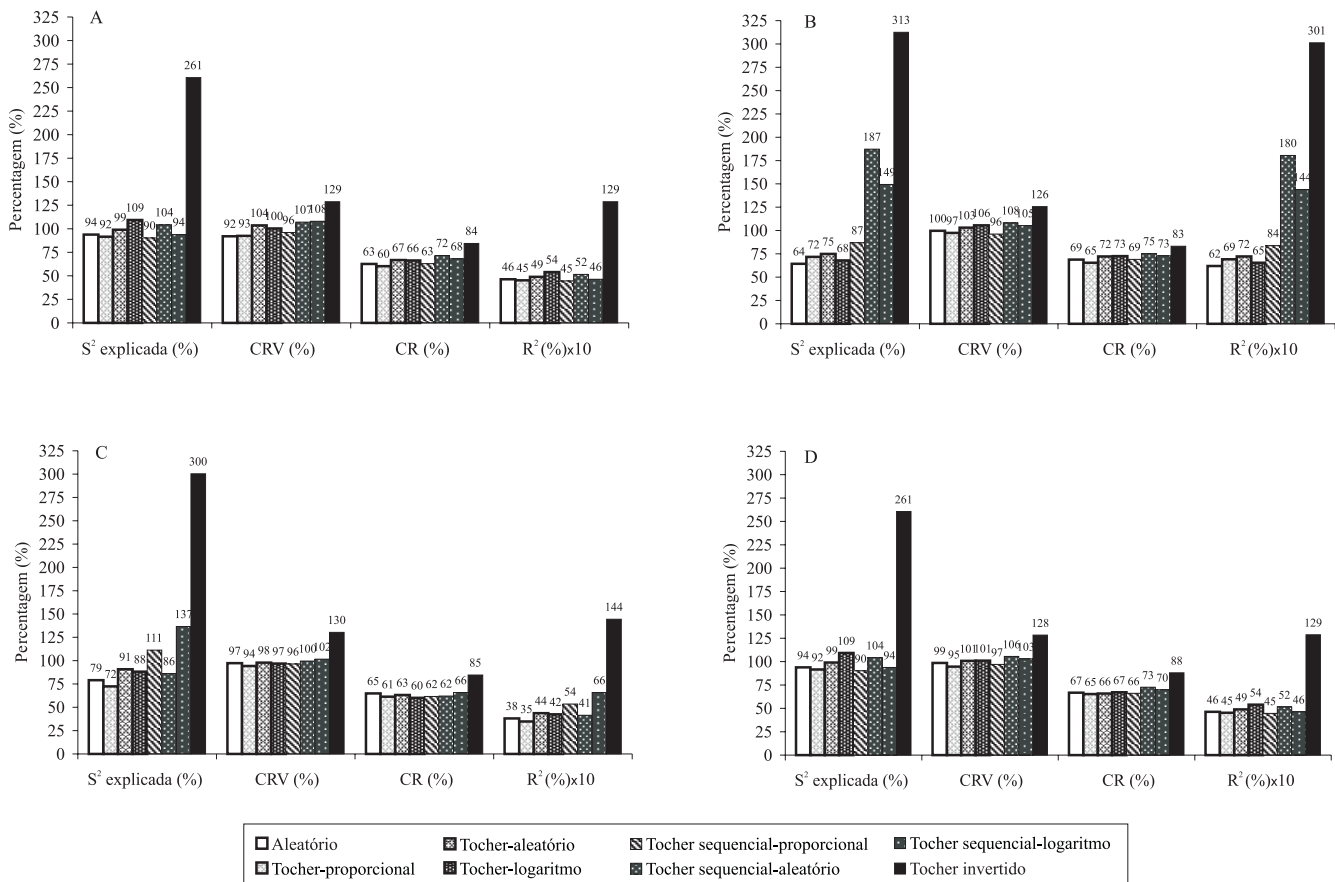
aglomeração de genótipos com maior divergência, há menor redução no valor do  $\sum_{i=1}^n x_i^2$ , o que leva a um acréscimo na variância e explica o grande número de variâncias diferentes (pelo teste F para variâncias populacionais, a 5% de probabilidade) entre a coleção nuclear e a coleção original (Tabela 2).

Hu et al. (2000) também fizeram uso dos testes t e F, para avaliar a diferença entre médias e variâncias entre a coleção nuclear e a coleção original, sem verificar diferenças significativas entre as coleções nucleares e as originais. Resultados semelhantes foram encontrados por Holbrook & Dong (2005), na obtenção de uma minicollection nuclear de amendoim, o que indica que as coleções nucleares são representativas da coleção original com relação à média e à variância.

Nas Figuras 1 e 2, estão apresentados os resultados da percentagem de variância da coleção original, explicada pela coleção nuclear ( $S^2$  explicada) do valor

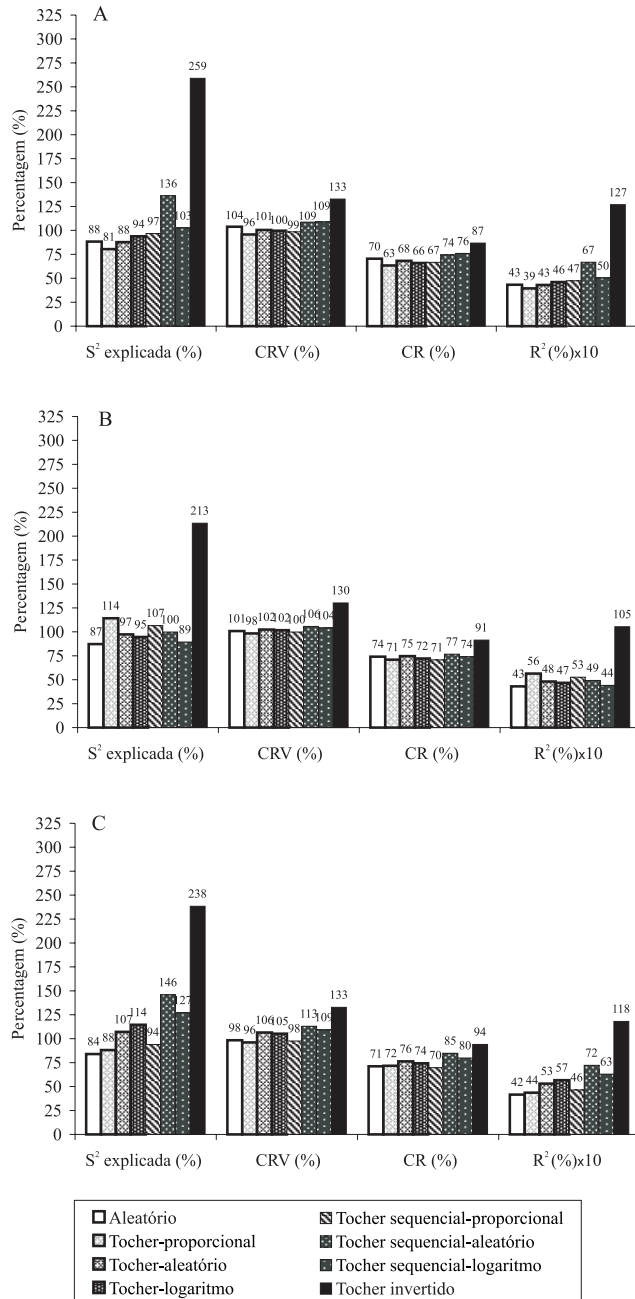
do coeficiente de determinação ( $R^2$ ), do coeficiente de similaridade (CR) e do coeficiente de similaridade da variância (CRV), entre as coleções nucleares e as coleções originais de diferentes tamanhos.

Na coleção original de 100 acessos, para o agrupamento de Tocher, as estratégias de amostragem constante e logarítmica foram superiores em mais de 30% à estratégia proporcional com base no valor da variância explicada ( $S^2$  explicada) e do  $R^2$  (Figura 3). Para o coeficiente de similaridade (CR) e de similaridade da variância (CRV), as estratégias aleatória e logarítmica foram superiores em 10% à estratégia proporcional. Nas coleções originais de 500, 750, 1.000 e 2.000 acessos, as coleções nucleares obtidas pelas estratégias de amostragem constante e logarítmica geraram estimativas de  $S^2$  explicadas,  $R^2$ , CR e CRV com valores relativamente superiores aos obtidos por amostragem proporcional.



**Figura 1.** Valores de variância da coleção original, explicada pela coleção nuclear ( $S^2$  explicada), coeficiente de similaridade da variância (CRV), coeficiente de similaridade entre a coleção original e sua coleção nuclear (CR) e coeficiente de determinação ( $R^2$ ) para populações de: A, 100; B, 250; C, 500; D, 750 acessos.

No método de agrupamento de Tocher sequencial, a estratégia de amostragem constante – em que se seleciona um acesso em cada grupo até se chegar ao número de acessos para formação da coleção nuclear – formou coleções nucleares com maiores estimativas

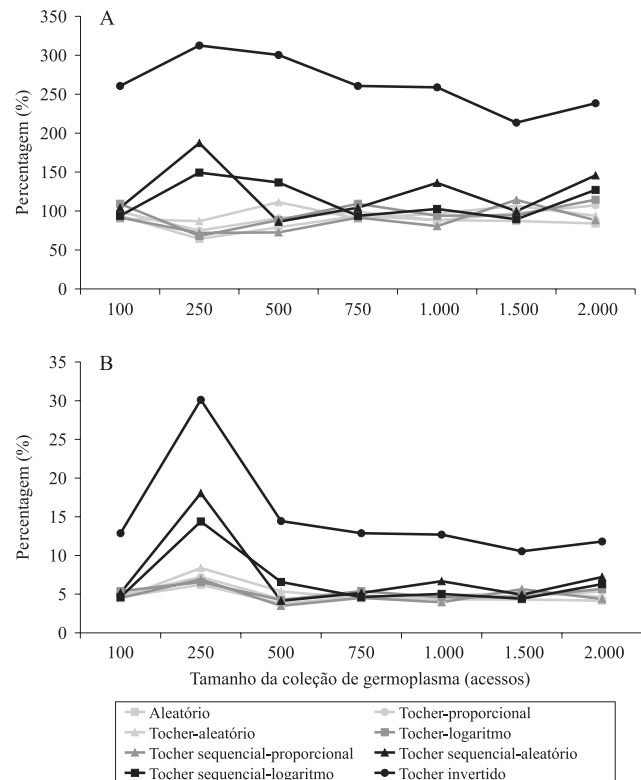


**Figura 2.** Valores de variância da coleção original, explicada pela coleção nuclear ( $S^2$  explicada), coeficiente de similaridade da variância (CRV), coeficiente de similaridade entre a coleção original e sua coleção nuclear (CR) e coeficiente de determinação ( $R^2$ ) para populações de: A, 1.000; B, 1.500; C, 2.000 acessos.

de  $S^2$  explicada,  $R^2$ , CR e CRV, em relação às demais estratégias de amostragem logarítmica e proporcional ao tamanho de grupo.

Van Hintum et al. (1995) avaliaram a melhor estratégia de amostragem dentro dos grupos, obtidos de uma coleção de 100 acessos, e verificaram poucas diferenças entre as amostragens constante, logarítmica e a proporcional. Contudo, eles citam que a estratégia de alocação constante foi menos efetiva que a proporcional e a logarítmica.

O resultado de Van Hintum et al. (1995) não foi confirmado no presente trabalho, uma vez que, para a população de 100 acessos, as estratégias de amostragem constante e logarítmica geraram melhores resultados do que a proporcional pelo agrupamento de Tocher, e a estratégia aleatória foi a que gerou coleções nucleares que melhor representaram as coleções originais, quando o agrupamento foi realizado pelo método de Tocher sequencial. Isso pode ter ocorrido em razão do número de grupos formados por apenas um genótipo quando foi empregado o agrupamento de Tocher. Assim, quando



**Figura 3.** Comportamento da fração da variância da coleção original explicada pela coleção nuclear ( $S^2$  explicada), e do coeficiente de determinação ( $R^2$ ) (A e B, respectivamente), em relação aos diferentes tamanhos das coleções originais.

se emprega a estratégia de amostragem aleatória entre os grupos, é possível selecionar genótipos de grupos com maiores distâncias entre si.

O método de amostragem logarítmico tem sido apresentado como um dos que propicia melhores resultados de representação da coleção original no estabelecimento de coleções nucleares, e é utilizado na obtenção de coleção nuclear de arroz (Abadie et al., 2005) e de sorgo (Grenier et al., 2001), entre outras. Contudo, com base em dados isoenzimáticos, Chandra et al. (2002) recomendam a utilização de amostragem proporcional ou aleatória para a obtenção da coleção nuclear de batatas.

A Figura 4 A contém os resultados de CR gerados pelas coleções nucleares obtidas por diferentes estratégias avaliadas. Os valores do coeficiente de similaridade tendem a aumentar com o aumento do número de acessos da coleção original e chegam a ser superiores a 80% na coleção nuclear, obtida pelo agrupamento de Tocher sequencial com estratégia de amostragem constante ou igual a 79,85%, no caso de amostragem logarítmica com o mesmo método de agrupamento. O agrupamento de Tocher sequencial foi, em geral, mais eficiente do que o de Tocher para a obtenção de coleções nucleares, com estimativas do coeficiente de similaridade superiores. Resultados semelhantes foram verificados para o coeficiente de similaridade da variância (Figura 4 B).

Foram verificados altos valores de  $S^2$  explicada e de  $R^2$ , geradas pela avaliação das coleções nucleares a partir da coleção original de 250 acessos, e foram menores para a coleção original de 500 acessos (Figura 4). Isto ocorreu porque na coleção de 100 e 250 acessos a amostragem foi de 10%, enquanto na de 500 acessos ou mais foi de 5%. Com relação aos valores de CR e de CRV, essa variação não foi tão acentuada.

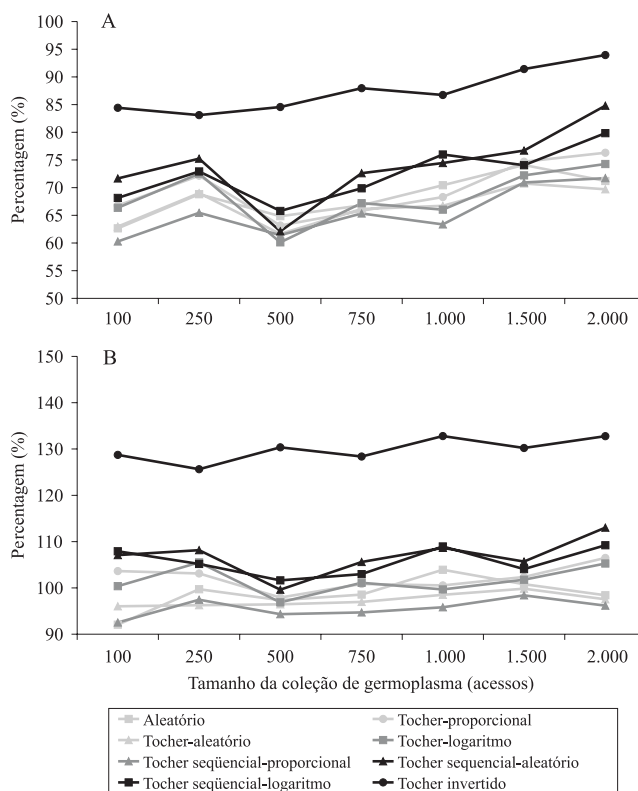
Valores do CRV acima de 100% foram verificados para a maioria das coleções nucleares, o que indica que a variância média delas é maior do que a da coleção original. Upadhyaya et al. (2002) também obtiveram valores de CRPD superiores a 100%, o que indica que a coleção nuclear representou a variação das características da coleção original.

O método de Tocher invertido propiciou estimativas maiores de  $S^2$  explicada,  $R^2$ , CR e CRV, e foi a melhor estratégia para obtenção de coleções nucleares, a partir de coleções originais com diferentes tamanhos. Essa estratégia foi a única com CR superior a 80% em todos os casos; o limite mínimo exigido na elaboração de

coleção nuclear para o coeficiente de coincidência entre coleção nuclear e coleção original é de 80% (Hu et al., 2000).

O tamanho da coleção original não afetou as estratégias de obtenção da coleção nuclear, seja daquelas que utilizam métodos de agrupamento seguido de amostragem ou não, e nenhuma estratégia de amostragem se mostrou relativamente superior para um dado tamanho de coleção original.

Este estudo demonstra que a estratégia de amostragem constante gera coleções nucleares com maior representação da coleção original do que a estratégia proporcional, quando se faz uso do método de Tocher e de Tocher sequencial para agrupamento. O agrupamento de Tocher sequencial possibilita aumento na eficácia de representação na coleção nuclear. Contudo, o método de Tocher com critério de aglomeração inverso, que não faz uso de técnicas de agrupamento ou estratificação, propiciou melhores



**Figura 4.** Comportamento do coeficiente de similaridade (CR; A) e do coeficiente de similaridade da variância (CRV, B) entre a coleção nuclear e a respectiva coleção original, geradas a partir das diferentes estratégias de obtenção de coleções nucleares avaliadas.

resultados de representação da coleção original pela coleção nuclear, para todos os tamanhos de coleção original avaliados.

### Conclusões

1. O agrupamento de Tocher sequencial possibilita obtenção de coleções nucleares mais representativas da coleção original.

2. A estratégia de amostragem constante gera coleções nucleares mais representativas da coleção original do que a estratégia proporcional e logarítmica, quando se faz uso do agrupamento de Tocher sequencial.

3. O método de Tocher com critério de agrupamento inverso, independentemente do tamanho da coleção original, propicia melhores resultados de representação da coleção original pela coleção nuclear.

4. O tamanho da coleção original não tem efeito sobre as estratégias de obtenção da coleção nuclear, sejam as que utilizem métodos de agrupamento seguido de amostragem ou não.

### Referências

- ABADIE, T.; CORDEIRO, C.M.T.; FONSECA, J.R.; ALVES, R. de B. das N.; BURLE, M.L.; BRONDANI, C.; RANGEL, P.H.N.; CASTRO, E. da M. de; SILVA, H.T. da; 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, p.129-136, 2005.
- 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, 1989. p.136-156.
- CHANDRA, S.; HUAMAN, Z.; HARI KRISHNA, S.; ORTIZ, R. Optimal sampling strategy and core collection size of Andean tetraploid potato based on isozyme data – a simulation study. **Theoretical and Applied Genetics**, v.104, p.1325-1334, 2002.
- CRUZ, C.D. **Programa GENES: biometria**. Viçosa: UFV, 2006. 385p.
- CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 2.ed. Viçosa: UFV, 2006. v.2, 585p.
- FRANCO, J.; CROSSA, J.; TABA, S.; SHANDS, H. A sampling strategy for conserving genetic diversity when forming core subsets. **Crop Science**, v.45, p.1035-1044, 2005.
- GHISLAIN, M.; ZHANG, D.; FAJARDO, D.; HUAMÁN, Z.; HIJMANS, R.J. Marker-assisted sampling of the cultivated Andean potato *Solanum phureja* collection using RAPD markers. **Genetic Resources and Crop Evolution**, v.46, p.547-555, 1999.
- GOUESNARD, B.; BATAILLON, T.M.; DECOUX, G.; ROZALE, C.; SCHOEN, D.J.; DAVID, J.L. MSTRAT: an algorithm for building germ plasm core collections by maximizing allelic or phenotypic richness. **Journal of Heredity**, v.92, p.93-94, 2001.
- GRENIER, C.; HAMON, P.; BRAMEL-COX, P.J. Core collection of sorghum: II. Comparison of three random sampling strategies. **Crop Science**, v.41, p.241-146, 2001.
- HOLBROOK, C.C.; DONG, W. Development and evaluation of a mini core collection for the U.S. peanut germoplasm collection. **Crop Science**, v.45, p.1540-1544, 2005.
- 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.
- HUAMÁN, Z.; AGUILAR, C.; ORTIZ, R. Selecting a Peruvian sweetpotato core collection on the basis of morphological, eco-geographical, and disease and pest reaction data. **Theoretical and Applied Genetics**, v.98, p.840-844, 1999.
- OLIVEIRA, A.C.B. de; SEDIYAMA, M.A.N.; PEDROSA, M.W.; GARCIA, N.C.P.; GARCIA, S.L.R. Divergência genética e descarte de variáveis em alface cultivada sob sistema hidropônico. **Acta Scientiarum. Agronomy**, v.26, p.211-217, 2004.
- SILVA, C.M.; GONÇALVES-VIDIGAL, M.C.; VIDIGAL FILHO, P.S.; SCAPIM, C.A.; DAROS, E.; SILVÉRIO, L. Genetic diversity among sugarcane clones (*Saccharum* spp.). **Acta Scientiarum. Agronomy**, v.27, p.315-319, 2005.
- SOUZA, F. de F.; QUEIRÓZ, M.A. de; DIAS, R. de S.C. Divergência genética em linhagens de melancia. **Horticultura Brasileira**, v.23, p.179-183, 2005.
- UPADHYAYA, H.D.; BRAMEL, P.; ORTIZ, R.; SINGH, S. Developing a mini core of peanut for utilization of genetic resources. **Crop Science**, v.42, p.2150-2156, 2002.
- VAN HINTUM, T.J.L.; VON BOTHMER, R.; VISSER, D.L. Sampling strategies for composing a core collection of cultivated barley (*Hordeum vulgare* s. lat) collected in China. **Hereditas**, v.122, p.7-17, 1995.
- VASCONCELOS, E.S. de; CRUZ, C.D.; BHERING, L.L.; FERREIRA, A. Estratégias de amostragem e estabelecimento de coleções nucleares. **Pesquisa Agropecuária Brasileira**, v.42, p.507-514, 2007a.
- VASCONCELOS, E.S. de; CRUZ, C.D.; BHERING, L.L.; RESENDE JÚNIOR, M.F.R. Método alternativo para análise de agrupamento. **Pesquisa Agropecuária Brasileira**, v.42, p.1421-1428, 2007b.

Recebido em 11 de maio de 2010 e aprovado em 17 de novembro de 2010