



Consequências da transformação de dados multialélicos em binários na análise de diversidade genética

ÉDER CRISTIAN MALTA DE LANES¹; JOSÉ MARCELO SORIANO VIANA²; GEÍSA PINHEIRO PAES³; MÁRCIA FABIANA BARBOSA DEPAULA⁴; CIRO MAIA⁵; EVELINE TEIXEIRA CAIXETA⁶; GLAUCO VIEIRA MIRANDA⁷; 1,2,3,4,5,7. UNIVERSIDADE FEDERAL DE VIÇOSA, VICOSA, MG, BRASIL; 6. PESQUISADORA EMBRAPA CAFÉ, VICOSA, MG, BRASIL; edercml@gmail.com

Resumo: Pesquisas em diversidade genética com base em marcadores codominantes tem utilizado a transformação como forma de ajustar os dados as medidas de similaridade ou distância genética apropriada para informações binárias. O objetivo desse trabalho foi avaliar as consequências da transformação de dados multialélicos em binários na análise de diversidade genética. Os três índices multialélicos foram concordantes na identificação do par de linhagens mais similares (78 e 79), mas não houve consenso em relação ao par mais divergente. Independente do método de agrupamento, o índice de Lynch proporcionou o maior número de grupos com pelo menos duas linhagens, associado à menor taxa de erro. A transformação de dados multialélicos em binários não se justifica pela existência de medida de similaridade e de dissimilaridade apropriadas e porque os critérios de transformação adotados conduziram a análises de diversidade com resultados de agrupamento distintos e com maior taxa de erro.

Palavras-chave: agrupamento, distância genética, microssatélites

Introdução

A caracterização molecular mede o nível de similaridade ou dissimilaridade genética entre os materiais, o que é importante para a manutenção de recursos genéticos e desenvolvimento de variedades (Romero et al., 2009). Em muitos estudos de diversidade com base em marcadores microssatélites (SSR) no melhoramento vegetal foi empregada medida de similaridade ou distância genética apropriada para dados binários (coeficiente de Jaccard, Sorensen-Dice, Coincidência Simples entre outros), evidenciando algum tipo de transformação. Nos estudos com linhagens ou linhas puras em que se empregou medida de dissimilaridade calculada com base nas frequências alélicas (distância de Roger e Modificada de Roger, etc) (Reif et al., 2005), o arquivo de frequências gênicas é binário, equivalendo à transformação de cada alelo em um loco dominante. A transformação, *a priori*, não se justifica, pois existe medida de similaridade e de dissimilaridade adequadas para marcas multialélicas (Lynch, 1990; Smouse e Peakall, 1999; Reif et al., 2005). O objetivo desse trabalho foi avaliar as consequências da transformação de dados multialélicos em binários na análise de diversidade genética.



Material e Métodos

Noventa linhagens de milho tropical do Programa Milho® da Universidade Federal de Viçosa (UFV), oriundas de híbridos comerciais, foram genotipadas em relação a 81 locos SSR. O DNA foi extraído como descrito por Doyle e Doyle (1990), com modificações. Na análise dos dados multialélicos foram utilizados o índice de similaridade proposto por Lynch (1990), as medidas de dissimilaridade não ponderada e ponderada propostas por Smouse e Peakall (1999). Para avaliar o efeito da transformação em dados binários foram adotados os seguintes critérios: (1) definição aleatória do alelo dominante; (2) definição do alelo mais frequente como o dominante; e (3) transformação de cada alelo em um loco dominante. Para o primeiro critério realizaram-se quatro definições aleatórias dos alelos dominantes. A medida de similaridade para os dados binários foi o coeficiente de Dice (D). As distâncias genéticas foram computadas como $1 - SG$ (Similaridade genética), sendo $SG = S_{xy}$ ou $D, d_{ij}/L$. Para avaliar a eficiência dos agrupamentos foi utilizada a taxa de erro aparente da análise discriminante com base no método não paramétrico do k vizinho mais próximo. Para as análises foram empregados os softwares NTSYS-pc 2.21, SAS 9.0, GenA1Ex 6.4 e GENES. As medidas de dissimilaridade ponderada de Smouse e Peakall e a correspondentes distâncias genéticas foram obtidas por um software desenvolvido em REALbasic 5.5.

Resultados e Discussão

O índice de Lynch proporcionou a maior média de distância genética (0.6198) e o índice de Smouse e Peakall não ponderado gerou o menor valor médio (0.1834). Os três índices para dados multialélicos foram concordantes na identificação do par de linhagens mais similares (78 e 79), mas não houve consenso em relação ao par mais divergente (Tabela 1). A análise dos dados transformados revelou equivalência dos métodos quanto às distâncias genéticas mínima, média e máxima, embora os maiores valores de distância tenham sido obtidos com o critério 3 (Tabela 1). Contudo, em geral houve pouca ou nenhuma concordância quanto à identificação dos pares de linhagens mais e menos similares.

Tabela 1. Distâncias genéticas máxima, mínima e média, e linhagens menos e mais divergentes, obtidas pelos índices de Lynch, Smouse e Peakall ponderado (SPw) e não ponderado (SPu), e Dice, após transformação em padrões binários pelos critérios 1 (alelo dominante escolhido ao acaso – C1; quatro simulações), 2 (alelo dominante o mais frequente – C2) e 3 (um loco dominante para cada alelo – C3).

Dist./Par	Lynch	SPw	SPu	C11	C12	C13	C14	C2	C3
Maximo	0.8167	0.3198	0.2051	1.0000	1.0000	0.9231	0.9259	0.7358	0.8167
Mínimo	0.0758	0.0932	0.0956	0.0345	0.0476	0.0698	0.0370	0.0566	0.0758



Média	0.6198	0.2210	0.1834	0.5837	0.5894	0.4887	0.5393	0.4208	0.6199
Similar	78, 79	78, 79	78, 79 33, 53	79, 80	69, 80	79, 80	69, 80	78, 79	78, 79
Divergente	27, 84	10, 29	10, 73 27, 84	33, 39	23, 33 33, 39	33, 53	24, 53	33, 53	27, 84

Em relação ao número de grupos com pelo menos duas linhagens, observou-se que os critérios 1 e 3 proporcionaram resultados opostos com agrupamento pelos métodos de Tocher e UPGMA: maior número de grupos por um método e menor número pelo outro (Tabela 2). Nos dois casos, contudo, a menor taxa de erro foi obtida com o critério 3. Segundo Kosman e Leonard (2003) os coeficientes binários de Dice, Jaccard e Coincidência Simples, entre outros, não são apropriados para diploides com marcadores codominantes. E independente do método de agrupamento, a análise de diversidade com base na medida de Lynch proporcionou maior número de grupos com pelo menos duas linhagens e menor taxa de erro (Tabela 2).

Tabela 2. Número de grupos e de linhagens isoladas obtidos pelos métodos de Tocher e UPGMA, empregando os índices de Lynch, Smouse e Peakall ponderado (SPw) e não ponderado (SPu), e Dice, após transformação em dados binários pelos critérios 1 (alelo dominante escolhido ao acaso – C1; quatro simulações), 2 (alelo dominante o mais frequente – C2) e 3 (um loco dominante para cada alelo – C3), e taxa de erro.

Método		Lynch	SPw	SPu	C11	C12	C13	C14	C2	C3
Tocher	Grupos	18	1	11	4	7	13	11	9	17
	Linhagens	8	9	11	5	3	11	6	9	10
	Taxa de erro	14.63	-	21.52	15.29	40.23	44.30	28.57	39.51	16.25
UPGMA	Grupos	18	1	2	7	10	14	15	9	3
	Linhagens	5	1	0	4	4	10	6	2	3
	Taxa de erro	8.23	-	11.11	36.05	34.88	16.25	22.62	51.14	10.34

Conclusão

A transformação de dados multialélicos em binários não se justifica pela existência de medida de similaridade e de dissimilaridade apropriadas e porque os critérios de transformação adotados conduziram a análises de diversidade com resultados de agrupamento distintos e com maior taxa de erro.

Agradecimentos



Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, pela bolsa concedida ao primeiro autor.

Referências Bibliográficas

- DOYLE, J. J. T.; DOYLE, J. L. Isolation of plant DNA from fresh tissue. **Focus**, 12, 13-15, 1990.
- KOSMAN, E., LEONARD, K.J. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. **Molecular Ecology**, 14: 415-424, 2005.
- LYNCH, M.L. The similarity index and DNA fingerprinting. **Molecular Biology and Evolution**, 7: 478-484, 1990.
- REIF, J.C., MELCHINGER, A.E., FRISCH, M. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. **Crop Science**, 45: 1-7, 2005.
- ROMERO, G.; ADEVA, C.; BATTAD, Z. Genetic fingerprinting: advancing the frontiers of crop biology research. **Philippine Science Letters**, 2:8–13, 2009.
- SMOUSE, P.E., PEAKALL, R. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. **Heredity**, 82: 561-573, 1999.