

Parecença entre Acessos de Feijoeiro Comum de Grãos Brancos Utilizando o Método de Harrison

Jaison Pereira de Oliveira¹
Maria José Del Peloso²
Orlando Peixoto de Moraes³
Leonardo Cunha Melo⁴

Introdução

O banco ativo de germoplasma (BAG) da Embrapa Arroz e Feijão praticamente só possui informações subjetivas provenientes das cadernetas de coletas. Modelos matemáticos relacionando a parença entre acessos, utilizando variáveis subjetivas, têm sido pouco explorados nos estudos de recursos genéticos. A parença é dividida em duas categorias, similaridade e dissimilaridade. Modelos que levam em consideração o par (0,0) podem ter o valor do coeficiente de parença enviesado devido ao elevado número de coincidências de zeros. Por outro lado, o termo parença mede o grau de parentesco entre dois indivíduos. A opção do uso de tal termo se deve ao fato de representar em uma única palavra os termos similaridade e dissimilaridade. Diante disso, o objetivo deste trabalho foi estudar o coeficiente de parença através da aplicação do modelo de Harrison, em variáveis qualitativas multicategóricas e determinar a matriz de

distâncias entre acessos tradicionais de feijoeiro comum visando detectar possíveis redundâncias entre acessos.

Material e Método

Foram utilizados 18 acessos tradicionais de feijoeiro comum de grãos brancos do BAG-Embrapa Arroz e Feijão. Os descritores, tipo de planta (HC), tamanho de grão (TG), unidade federativa (UF), classe de solo (CS) e classe de altitude (CA), foram transformados em variáveis binárias através da criação de variáveis fictícias (Tabela 1). Nesse caso, seja y' uma variável qualitativa, formada por um vetor de variáveis qualitativas nominais: $y' = (y_1, y_2, \dots, y_l)$ onde a i -ésima componente assume l_i níveis, codificados de modo que $y_j = j$, com $j = 1, 2, \dots, l_i$. Ao transformar essa variável em uma variável binária, cada componente (critério qualitativo) contribuirá para a geração de l_i variáveis binárias $x_k (i)$ tal que:

¹ Engenheiro agrônomo, Doutor em Genética e Melhoramento de Plantas, pesquisador da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, jaison@cnpaf.embrapa.br

² Engenheira agrônoma, Doutora em Genética e Melhoramento de Plantas, pesquisadora da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, mjpeloso@cnpaf.embrapa.br

³ Engenheiro agrônomo, Doutor em Genética e Melhoramento de Plantas, pesquisador da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, peixoto@cnpaf.embrapa.br

⁴ Engenheiro agrônomo, Doutor em Genética e Melhoramento de Plantas, pesquisador da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, leonardo@cnpaf.embrapa.br

$$x_i(i) = \begin{cases} 1 & \text{se } y_i = k \\ 0 & \text{caso contrário} \end{cases}$$

Assim, o vetor y de dimensão n é transformado no vetor x de dimensão p , formado apenas por componentes binários, e por consequência, y' se transforma em x' :

$$y' = (y_1, y_2, \dots, y_l) \rightarrow x' = (\underbrace{0, \dots, 1, \dots, 0}_{l_1}, \dots, \underbrace{0, \dots, 1, \dots, 0}_{l_i})$$

Os vetores de variáveis binárias permitem gerar uma tabela de dupla entrada. Essa tabela possibilita estimar o coeficiente de parença entre dois objetos (acessos, genótipos, etc.). Condensando os valores binários em uma tabela de dupla entrada tem-se:

	objeto 1 (o_1)		
	1	0	
objeto 2 (o_2)			
1	a	b	
0	c	d	

Onde, a : representa o par (1,1), e corresponde ao total de concordâncias positivas ocorridas entre os dois objetos; b : representa o par (0,1), e corresponde ao total de discordâncias ocorridas entre o objeto 1 e o objeto 2, respectivamente; c : representa o par (1,0), e corresponde ao total de discordâncias ocorridas entre o objeto 1 e o objeto 2, respectivamente; e d : representa o par (0,0), e corresponde ao total de concordâncias negativas ocorridas entre os dois objetos.

A partir dos valores gerados pela tabela de dupla entrada (a , b , c , d), estima-se o coeficiente de parença. Vários coeficientes de parença podem ser encontrados em Gower (1985) e Bussab et al. (1990). Aqui se fez opção por usar o coeficiente proposto por Harrison (1968). Tal opção baseia-se no fato de que alguns coeficientes não são aplicáveis quando suas variáveis geram uma indefinição matemática por conta de valores nulos. Por outro lado, coeficientes que levam em consideração o par (0,0) representado por d , na tabela de dupla entrada, podem ter o valor do coeficiente de parença poluído (enviesado) devido ao elevado

número de coincidências de zeros, já esperado por conta do tipo de transformação. Alguns coeficientes não seriam adequados para variáveis binárias oriundas de dados qualitativos de mais de dois níveis.

Como o número de níveis de cada variável possui heterogeneidade (p. ex. quatro níveis para tamanho de semente, cinco para tipo de planta, etc.), no coeficiente de Harrison (s), cada variável é ponderada pelo número de níveis que possui atenuada pelo cálculo do seu logaritmo. O modelo matemático de Harrison é:

$$s(o_1, o_2) = \frac{\sum_{i=1}^l \ln l_i}{\sum_{i=1}^l \ln l_i} I(y_i(o_1), y_i(o_2)) = \sum_{i=1}^l w_i I(y_i(o_1), y_i(o_2)) \quad \text{com} \quad w_i = \ln l_i / \sum_{i=1}^l \ln l_i$$

Onde a função I é indicadora de coincidência de níveis, isto é,

$$I(y_i(o_1), y_i(o_2)) = \begin{cases} 1 & \text{se } y_i(o_1) = y_i(o_2) \\ 0 & \text{se } y_i(o_1) \neq y_i(o_2) \end{cases}$$

Resultados e Discussão

As frequências relativas de similaridade (w_i) encontradas nos descritores foram de 23,0% para descritores com sete níveis (CS e AL), 19,0% para cinco níveis (HC e UF) e 16,0% para quatro níveis (TG) (Tabela 1). Essa análise mostrou que descritores com muitos níveis contribuem com maior frequência relativa para a similaridade no caso de coincidência entre dois acessos. A parença entre os acessos estudados pode ser observada na Tabela 2. Os acessos 9 e 10 são redundantes em relação aos descritores utilizados, apresentando valor de similaridade igual a 1,00 (100%). Altas similaridades foram observadas entre os acessos 11 e 15 (0,84), 14 e 16 (0,84), 12 com o 14 e com o 15 (0,81) e 14 e 15 (0,81). Os acessos 1, 2 e 3, 2 e 4, 8 com o 9 e com o 10, apresentaram similaridade de 0,77. Por outro lado, a maior divergência foi observada no acesso 5, variando de 0,65 a 1,00 de dissimilaridade. O acesso 5 apresentou dissimilaridade máxima (1,00) com onze acessos. O método de Harrison possibilitou detectar acessos redundantes, sendo uma ferramenta de decisão e definição no descarte de genótipos iguais, bem como seleção de acessos divergentes.

Tabela 1. Matriz de variáveis morfogeográficas e variáveis fictícias entre acessos tradicionais em feijoeiro comum de grão branco.

Acessos	Variáveis morfogeográficas																											
	Tipo de planta				Tamanho de grão				Unidade Federativa				Classe de solo								Altitude							
	H1	H2	H3	H4	H6	P	M	G	MG	AL	PE	MG	PR	RG	S1	S2	S3	S4	S5	S6	S7	A1	A2	A3	A4	A5	A6	A7
1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
2	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
3	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
4	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
5	0	1	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
6	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
7	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
8	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0
9	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
10	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
11	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0
12	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0
13	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0
14	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0
15	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0
16	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0
17	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0
18	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Total	4	2	8	2	2	4	9	4	1	4	1	1	1	11	4	6	3	2	1	1	1	2	3	2	6	3	1	1
w _i	0,19				0,16				0,19				0,23								0,23							

H1: Determinado tipo I, H2: Indeterminado tipo II, H3: Indeterminado tipo III, H4: Indeterminado tipo IV, H6: Indeterminado entre III e IV; P: pequena (< 21 g), M: média (21 g a 30 g), G: grande (31 g a 40 g) e MG: muito grande (> 40 g) em relação a massa de 100 grãos; AL: Alagoas, PE: Pernambuco, MG: Minas Gerais, PR: Paraná, RG: Rio Grande do Sul; S1: Argilossolo Vermelho-Amarelo, S2: Cambissolo Hápico, S3: Latossolo Vermelho-Amarelo, S4: Neossolo Litólico, S5: Neossolo Regolítico, S6: Nitossolo Vermelho, S7: Planossolo Hápico; A1: 0 a 100, A2: 200 a 300, A3: 300 a 400, A4: 500 a 600, A5: 600 a 700, A6: 700 a 800 e A7: 900 a 1000 (em relação a altitude em metros).

Tabela 2. Matriz de distâncias da parecença (similaridade e dissimilaridade, diagonal inferior e superior, respectivamente) entre acessos tradicionais em feijoeiro comum de grão branco, pelo método de Harrison.

		Acessos tradicionais de feijoeiro comum de grãos brancos																	
Acessos		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Acessos tradicionais de feijoeiro comum de grãos brancos	1	·	0,23	0,23	0,46	1,00	0,65	1,00	0,77	1,00	1,00	1,00	0,65	0,81	0,84	0,84	1,00	1,00	0,42
	2	0,77	·	0,23	0,23	1,00	0,65	1,00	1,00	1,00	1,00	1,00	0,65	0,81	0,84	0,84	1,00	1,00	0,65
	3	0,77	0,77	·	0,46	1,00	0,65	0,77	1,00	1,00	1,00	1,00	0,65	0,58	0,84	0,84	1,00	1,00	0,65
	4	0,54	0,77	0,54	·	1,00	0,65	1,00	1,00	1,00	1,00	1,00	0,65	0,81	0,84	0,84	1,00	1,00	0,65
	5	0,00	0,00	0,00	0,00	·	1,00	0,84	0,77	0,77	0,77	0,84	1,00	1,00	1,00	1,00	1,00	0,65	1,00
	6	0,35	0,35	0,35	0,35	0,00	·	1,00	1,00	1,00	1,00	0,77	0,42	0,81	0,61	0,61	0,77	1,00	0,65
	7	0,00	0,00	0,23	0,00	0,16	0,00	·	0,81	0,81	0,81	0,84	1,00	0,77	1,00	1,00	1,00	0,61	1,00
	8	0,23	0,00	0,00	0,00	0,23	0,00	0,19	·	0,23	0,23	0,81	0,81	0,65	0,81	0,81	0,81	0,81	0,58
	9	0,00	0,00	0,00	0,00	0,23	0,00	0,19	0,77	·	0,00	0,81	0,81	0,65	0,81	0,81	0,81	0,81	0,81
	10	0,00	0,00	0,00	0,00	0,23	0,00	0,19	0,77	1,00	·	0,81	0,81	0,65	0,81	0,81	0,81	0,81	0,81
	11	0,00	0,00	0,00	0,00	0,16	0,23	0,16	0,19	0,19	0,19	·	0,35	0,58	0,35	0,16	0,35	0,65	0,81
	12	0,35	0,35	0,35	0,35	0,00	0,58	0,00	0,19	0,19	0,19	0,65	·	0,39	0,19	0,19	0,35	0,81	0,46
	13	0,19	0,19	0,42	0,19	0,00	0,19	0,23	0,35	0,35	0,35	0,42	0,61	·	0,58	0,58	0,58	0,81	0,62
	14	0,16	0,16	0,16	0,16	0,00	0,39	0,00	0,19	0,19	0,19	0,65	0,81	0,42	·	0,19	0,16	0,81	0,65
	15	0,16	0,16	0,16	0,16	0,00	0,39	0,00	0,19	0,19	0,19	0,84	0,81	0,42	0,81	·	0,35	0,81	0,65
	16	0,00	0,00	0,00	0,00	0,00	0,23	0,00	0,19	0,19	0,19	0,65	0,65	0,42	0,84	0,65	·	0,81	0,81
	17	0,00	0,00	0,00	0,00	0,35	0,00	0,39	0,19	0,19	0,19	0,35	0,19	0,19	0,19	0,19	0,19	·	0,81
	18	0,58	0,35	0,35	0,35	0,00	0,35	0,00	0,42	0,19	0,19	0,19	0,54	0,38	0,35	0,35	0,19	0,19	·

Referências

BUSSAB, W. de O.; MIAZAKI, E. S.; ANDRADE, D. F. de. **Introdução à análise de agrupamentos**. São Paulo: USP-IME, 1990. 105 p. Trabalho apresentado no 9. Simpósio Nacional de Probabilidade e Estatística, São Paulo, julho de 1990.

GOWER, J. C. Measures of similarity, dissimilarity, and distance. In: KOTZ, S.; JOHNSON, N. L.; READ, C. B. (Ed.). **Encyclopedia of statistical sciences**. New York: J. Wiley, 1985. v. 5. p. 397-405.

HARRISON, P.J. A method of cluster analysis and some applications. **Journal of the Royal Statistical Society. Series C. Applied Statistics**, London, v. 17, n. 3, p. 226-236, 1968.

Comunicado Técnico, 181



Ministério da Agricultura,
Pecuária e Abastecimento



Exemplares desta edição podem ser adquiridos na:

Embrapa Arroz e Feijão

Rodovia GO 462 Km 12 Zona Rural

Caixa Postal 179

75375-000 Santo Antônio de Goiás, GO

Fone: (62) 3533 2194

Fax: (62) 3533 2100

E-mail: sac@cnpaf.embrapa.br

1ª edição

1ª impressão (2009): 1.000 exemplares

Comitê de publicações

Presidente: Luís Fernando Stone

Secretário-Executivo: Luiz Roberto R. da Silva

Membro: Joaquim Geraldo Cápio da Silva

Expediente

Supervisor editorial: Camilla Souza de Oliveira

Revisão de texto: Camilla Souza de Oliveira

Normalização bibliográfica: Ana Lúcia D. de Faria

Tratamento das ilustrações: Fabiano Severino

Editoração eletrônica: Fabiano Severino