

Parecença Entre Acessos De Feijoeiro Comum De Grãos Vermelhos Utilizando O Método De Harrison

Jaison Pereira de Oliveira¹, Eduardo José Pinheiro², Maria José Del Peloso¹, Leonardo Cunha Melo¹ e Gabriel Bisinotto Pereira³.

Resumo

O termo parecença 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. Neste experimento foram utilizados 15 acessos tradicionais de feijoeiro comum de grãos vermelhos do BAG-Embrapa Arroz e Feijão. Fez-se a 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. O acesso um foi o que mais apresentou similaridade. Por outro lado, vários acessos apresentaram divergência máxima, ou seja, 1,00 de dissimilaridade. O acesso três apresentou dissimilaridade máxima (1,00) com oito acessos.

Introdução

O banco ativo de germoplasma (BAG) da Embrapa Arroz e Feijão praticamente só possuem informações subjetivas provenientes das cadernetas de coletas. Modelos matemáticos relacionando a parecença entre acessos, utilizando variáveis subjetivas, têm sido pouco explorados nos estudos de recursos genéticos. A parecença é dividida em duas categorias, similaridade e dissimilaridade. Modelos que levam em consideração o par (0,0) podem ter o valor do coeficiente de parecença enviesado devido ao elevado número de coincidências de zeros. Por outro lado, o termo parecença 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 parecença 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 15 acessos tradicionais de feijoeiro comum de grãos vermelhos 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).

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

Acesso	Tipo de Planta				Tamanho de grão			Unidade Federativa				Classe de solo				Altitude									
	H1	H2	H3	H5	P	M	G	M	ES	M	SP	RS	SC	S1	S2	S3	S4	A1	A2	A3	A4	A5	A6	A7	A8
1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
2	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
3	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
4	0	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
5	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
6	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
7	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0

¹: Pesquisador A Embrapa Arroz e Feijão, E-mail: jaison@cnpaf.embrapa.br; ²: Graduando em Agronomia, Universidade Federal de Goiás, E-mail: eduardoufg10@gmail.com; ³: Graduando em Agronomia, Universidade Federal de Goiás, E-mail: gabrielufg@gmail.com.

8	1	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1		
9	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	
10	1	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	
11	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	
12	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	
13	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	
14	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	
15	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	
Total	5	3	5	2	2	7	3	3	4	5	1	4	1	3	4	4	4	2	1	1	5	1	3	1	1
w_i	17,67				17,67				20,50				17,67				26,49								

H1: Determinado tipo I, H2: Indeterminado tipo II, H3: Indeterminado tipo III, H5: Indeterminado tipo II E III,; P: pequena (<21g), M: média (21g a 30g), G: grande (31g a 40g) e MG: muito grande (> 40g) em relação à massa de 100 grãos; ES: Espírito Santo, MG: Minas Gerais, SP: São Paulo, RG: Rio Grande do Sul, SC: Santa Catarina; S1: Argilossolo Vermelho, S2: Argilossolo vermelho-amarelo, S3: Cambissolo háplico, S4: Latossolo vermelho-amarelo; A1: 0 a 100, A2: 200 a 300, A3: 300 a 400, A4: 600 a 700, A5: 700 a 800, A6: 800 a 900; A7: 900 a 1000 e A8: > 1000 (em relação à Altitude em metros).

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 $y_i = j$, 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:

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

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' :

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

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:

	<i>objeto 1 (o₁)</i>		
	1	0	
1	<i>a</i>	<i>b</i>	
	<i>objeto 2 (o₂)</i>		
0	<i>c</i>	<i>d</i>	

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 um e o objeto dois, respectivamente; c : representa o par (1,0), e corresponde ao total de discordâncias ocorridas entre o objeto um e o objeto dois, 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 Bussab et al. (1990) e Gower (1985). 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.e. quatro níveis para tamanho de semente, quatro 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:

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

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 Discussões

As frequências relativas de similaridade (w_i) encontradas nos descritores foram de 26,5% para descritores com oito níveis (CA), 20,5% para cinco níveis (UF) e 17,7% para quatro níveis (HC, TG e CS) (Tabela 1). Esta 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 três e quatro, 11 e 12 e um e cinco apresentaram os maiores valores de similaridade (0,65, 0,65 e 0,63, respectivamente). Similaridades moderada foram observadas entre os acessos um e seis (0,53) e cinco e seis (0,56).

Tabela 2. Matriz de distâncias da parença entre acessos tradicionais em feijoeiro comum de grão vermelho, pelo método de Harrison.

		Acessos tradicionais de feijoeiro comum de grãos Vermelhos														
Acessos		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
		Distâncias de dissimilaridade entre acessos (diagonal superior)														
Acessos tradicionais de feijoeiro comum de grãos Vermelhos	1	-	0,53	0,62	0,62	0,38	0,47	1,00	1,00	1,00	1,00	0,56	0,56	0,65	0,65	1,00
	2	0,47	-	0,79	0,62	0,56	1,00	0,82	0,65	1,00	0,82	0,74	0,74	1,00	1,00	1,00
	3	0,38	0,21	-	0,35	0,82	0,82	1,00	1,00	1,00	1,00	1,00	0,82	1,00	1,00	0,82
	4	0,38	0,38	0,65	-	0,82	0,82	1,00	0,82	1,00	1,00	0,82	1,00	0,82	0,82	1,00
	5	0,62	0,44	0,18	0,18	-	0,44	0,62	0,79	0,79	1,00	0,56	0,74	0,82	0,82	1,00
	6	0,53	0,00	0,18	0,18	0,56	-	0,79	0,79	0,79	1,00	0,82	0,82	0,65	0,65	1,00
	7	0,00	0,18	0,00	0,00	0,38	0,21	-	0,79	0,62	1,00	0,82	0,82	1,00	0,56	0,74
	8	0,00	0,35	0,00	0,18	0,21	0,21	0,21	-	0,62	0,65	0,82	1,00	1,00	1,00	0,82
	9	0,00	0,00	0,00	0,00	0,21	0,21	0,38	0,38	-	0,65	0,82	1,00	0,82	1,00	0,65
	10	0,00	0,18	0,00	0,00	0,00	0,00	0,00	0,35	0,35	-	0,62	0,79	0,62	1,00	0,65
	11	0,44	0,26	0,00	0,18	0,44	0,18	0,18	0,18	0,18	0,38	-	0,35	0,62	0,65	0,82
	12	0,44	0,26	0,18	0,00	0,26	0,18	0,18	0,00	0,00	0,21	0,65	-	0,62	0,65	0,82
	13	0,35	0,00	0,00	0,18	0,18	0,35	0,00	0,00	0,18	0,38	0,38	0,38	-	0,65	0,82
	14	0,35	0,00	0,00	0,18	0,18	0,35	0,44	0,00	0,00	0,00	0,35	0,35	0,35	-	0,74

Alguns acessos apresentaram similaridade de 0,44. Tal valor equivale a 44,0% de parentesco entre acessos. O acesso um foi o que mais similaridade apresentou. Por outro lado, vários acessos apresentaram divergência máxima, ou seja, 1,00 de dissimilaridade. O acesso três apresentou dissimilaridade máxima (1,00) com oito acessos. O método de Harrison possibilitou detectar acessos com alto grau de parentesco sendo uma ferramenta de decisão e definição no descarte de genótipos iguais, bem como seleção de acessos divergentes.

Referências bibliográficas

BUSSAB, W. de O.; MIAZAKI, E.S.; ANDRADE, D.F. de **Introdução à análise de agrupamentos**. São Paulo: USP-IME, 1990. 105p. SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 9., São Paulo, 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: John Wiley & Sons, 1985. p.397-405. v.5.

HARRISON, P.J. A method of cluster analysis and some applications. **Applied Statistics**, v.17, n.3, p.226-236, 1968.

