

Pareceza entre Acessos Tradicionais de Feijão Jalo Utilizando o Método de Ward

Jaison Pereira de Oliveira¹, Leonardo Rodrigues Barros², Maria José Del Peloso¹, Leonardo Cunha Melo¹,
Silvando Carlos da Silva¹

INTRODUÇÃO

A informação quanto à variabilidade e a divergência genética dentro de uma espécie são essenciais para o uso racional dos recursos genéticos. A variabilidade genética existente no germoplasma tradicional de feijão (*Phaseolus vulgaris* L.) do Banco Ativo de Germoplasma - BAG, da Embrapa Arroz e Feijão têm sido plenamente reconhecidos. Esta variabilidade, existente em acessos tradicionais, é de fundamental importância nas estratégias econômica e social do desenvolvimento sustentável da cultura, pois o melhoramento genético do feijoeiro comum requer uma aquisição contínua de alelos, dependendo, portanto da diversidade presente nos recursos genéticos.

No estudo da diversidade genética em coleções de germoplasma podem-se utilizar caracteres morfológicos e ecogeográficos de natureza qualitativa, por meio de vários métodos, cuja escolha baseia-se na precisão desejada, na facilidade de análise e na forma como esses dados foram obtidos. As técnicas de análise multivariada podem ser utilizadas para avaliar a divergência entre indivíduos e detectar possíveis parecesas. Entre as técnicas da estatística multivariada, destacam-se a análise de componentes principais e os métodos de agrupamento.

A análise de agrupamento tem por finalidade reunir, por meio de critério de classificação, os acessos em grupos, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre grupos, dividindo assim, um grupo original de observações em vários grupos, segundo o critério de parecesas. O método de Ward foi escolhido para minimizar as diferenças internas de grupos e para evitar problemas com “encadeamento” das observações encontrados no método de ligação individual. Por outro lado, com a utilização do método de Ward, podem-se obter agrupamentos com regiões bem definidas, possibilitando a observação de proximidades de acessos com características semelhantes. A utilização desse método possibilita mostrar como se comportaram os dados, ou seja, a existência de algum padrão ou grau de organização dentro do conjunto de acessos estudados. Diante disto, o objetivo deste trabalho foi caracterizar a parecesas em 42 acessos tradicionais de feijão do grupo comercial jalo, reunindo-os em grupos de similaridade por meio da técnica de agrupamento de Ward aplicada a caracteres morfoecogeográficos.

MATERIAL E MÉTODOS

A base de dados de 42 acessos tradicionais de feijão do grupo comercial jalo da coleção de coleta de germoplasma tradicional da Embrapa Arroz e Feijão contém informações disponíveis compostas por dois descritores morfológicos (hábito de crescimento e tamanho de semente) e quatro descritores ecogeográfico (região geográfica, estado da federação, classe de solo e altitude). Foram agrupados os vetores de dados de forma a obter informações conclusivas sobre os acessos, com base de dados constituída por variáveis qualitativa e suas características.

Quando uma variável qualitativa possui mais de dois níveis é possível transformá-la em variáveis binárias por meio da criação de variáveis fictícias (BUSSAB et al., 1990). 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: $x_k(i) = 1$ se $y_i = k$ ou $x_k(i) = 0$. 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' : $y' = (y_1, y_2, \dots, y_l) \rightarrow x' = (0, \dots, 1, \dots, 0; \dots; 0, \dots, 1, \dots, 0)$.

Procurou-se agrupar os acessos por meio do Método de Ward (WARD, 1963) que constitui um método de agrupamento de dados que forma grupos de maneira a atingir sempre o menor erro interno entre os vetores que compõem cada grupo e o vetor médio do grupo. Isto equivale a buscar o mínimo desvio padrão entre os

¹Embrapa Arroz e Feijão, Caixa Postal 179, CEP 75375-000, Santo Antônio de Goiás, GO. E-mail: jaison@cnpaf.embrapa.br, ² Aluno de graduação em agronomia da Universidade Estadual de Goiás E-mail: leonardoagro92@hotmail.com

dados de cada grupo. No método de Ward, os grupos de dados são formados em etapas. No princípio, têm-se m grupos, ou seja, um grupo para cada vetor componente da base de dados. Neste estágio inicial o erro interno é nulo para todos os grupos, pois cada vetor que compõe cada grupo é o próprio vetor médio do grupo. Igualmente o desvio padrão para cada grupo é nulo. Na etapa subsequente, cada possibilidade de aglutinação entre os grupos 2 a 2 é verificada, e é escolhido o agrupamento que causa o menor aumento no erro interno do grupo. São $m \times m$ verificações, e a cada iteração tem-se $m - i$ grupos ($i =$ número de interações).

Com relação às informações, espera-se que a base de dados não descreva completamente os acessos a serem agrupados, isto principalmente porque cada característica tem valores 1 ou 0 somente; a ausência de valores intermediários entre 1 e 0 limita a capacidade da base de dados de descrever o conjunto de acessos. Também se sabe que nem todas as características presentes nos acessos estão presentes na base de dados, o que dificulta a tarefa, de distinguir entre um acesso e outro, particularidades genéticas.

RESULTADOS E DISCUSSÃO

Na aplicação do método de Ward, foram obtidas 41 etapas de agrupamento (devido aos 42 acessos). Tais etapas levaram a formação de um dendrograma com arranjos distintos para diversas distâncias de agrupamento (Figura 1) onde sete grupos foram sugeridos pelo gráfico do coeficiente de aglomeração (Figura 2). Observando-se os grupos formados, notou-se que os acessos reunidos nos últimos passos de agrupamento, cujos valores do coeficiente de aglomeração foram zero (Tabela 1) foram semelhantes entre si, indicando possíveis redundâncias. Essas possíveis redundâncias foram observadas em cinco grupos. Nesse caso, o agrupamento três, com cinco acessos apresentou redundância em dois pares.

Os agrupamentos cinco e seis, com nove acessos, formaram os grupos com maiores números de acessos, e ambos apresentaram pares de acessos redundantes. Tais grupos foram representados por acessos provenientes de Minas Gerais. O grupo quatro, com um acesso proveniente do Rio Grande do Sul, foi o mais divergente. O grupo 7 com quatro acessos, foi representado por três acessos de Minas Gerais e um acesso do Espírito Santo. Outro ponto semelhante está no grupo dois com seis acessos de Goiás e um acesso de São Paulo. Isso implica que os 42 acessos de feijão jalo podem ser resumidos a 34 acessos, sendo 7 oriundos dos agrupamentos mais 27 acessos individuais, assegurando a conservação de alelos distintos presentes no grupo. Essa conclusão é geral e se fundamentam em um modelo teórico que adota, a partir das variáveis estudadas, suposições sobre alguns parâmetros utilizados no cálculo das similaridades de acessos.

Para que o método de Ward atinja seu objetivo principal de formar um agrupamento coerente dos vetores de dados da base, é preciso adotar um critério que pode ser um número esperado de grupos, ou a configuração de menor dispersão geral, lembrando que, no princípio tinha-se 42 acessos dispersos nos 41 agrupamentos formados. Conforme se avança nas interações do método, agrupando os acessos, obtêm-se diferentes composições de grupos, até a união total. Assim, podem-se explorar todas as combinações possíveis dos dados de forma hierárquica aglomerativa, e escolher a mais interessante em conjunto com a dispersão dos elementos dentro de cada formação.

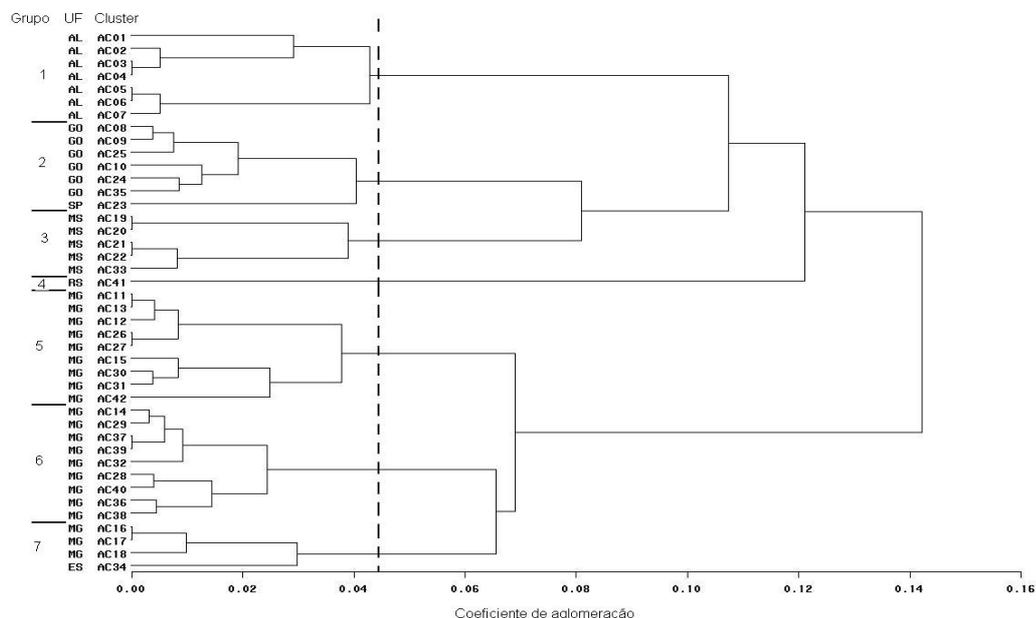


Figura 1. Agrupamento de 42 acessos de cultivares de feijão do grupo comercial jalo, pelo método de Ward.

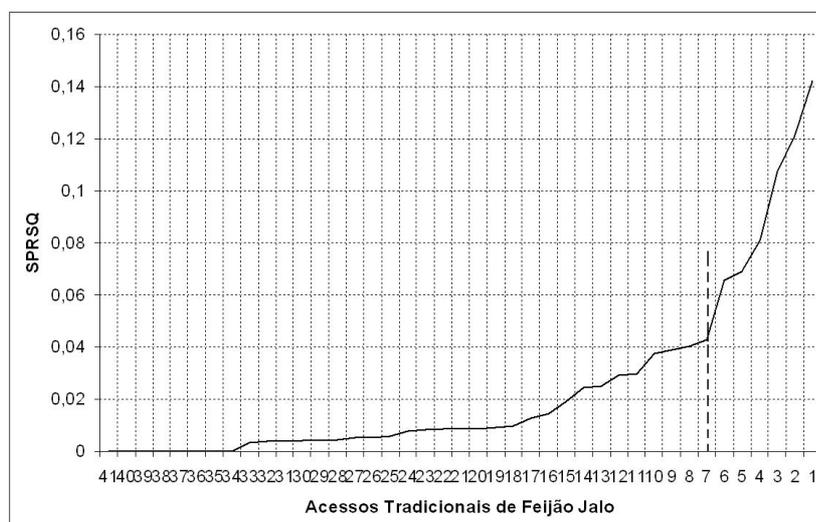


Figura 2. Distribuição dos Coeficientes de aglomeração (SPRSQ: correlação semiparcial ao quadrado) em relação aos passos do algoritmo de agrupamento pelo método de Ward, aplicado aos 42 acessos de feijão do grupo comercial jalo.

O método de Ward de agrupamento de dados propiciou uma visão mais clara dos resultados finais e da convergência do método para resultados reais. Verificou-se por meio de um acompanhamento da evolução do método, que os agrupamentos ocorrem da maneira esperada durante o desenvolvimento das interações. Foi possível ilustrar a influência da base de dados no resultado final. Além disso, ficou clara a relação entre as variáveis utilizadas e os grupos formados pelo método de Ward, sendo importantes para que se possam inferir resultados mais concretos e precisos a respeito de uma análise realizada para se determinar a parença entre acessos. Isto porque o método de Ward separa os dados em grupos bem definidos de modo a minimizar as diferenças interna de grupo. Tal método pode ser utilizado, quando se deseja compreender o padrão de composição de uma massa de dados contida em bancos de germoplasma.

Tabela 1 – Esquema de aglomeração de análise hierárquica dos 41 passos do agrupamento de 42 acessos tradicionais de feijão do grupo comercial jalo, pelo método de Ward.

Estágio	Agrupamento combinado		Coeficiente de aglomeração	Estágio	Agrupamento combinado		Coeficiente de aglomeração
	Acesso/agrupamento				Acesso/agrupamento		
41	AC03	AC04	0	20	AC24	AC35	0,009
40	AC05	AC06	0	19	CL25	AC32	0,009
39	AC11	AC13	0	18	CL38	AC18	0,010
38	AC16	AC17	0	17	AC10	CL20	0,013
37	AC19	AC20	0	16	CL30	CL28	0,014
36	AC21	AC22	0	15	CL24	CL17	0,019
35	AC26	AC27	0	14	CL19	CL16	0,024
34	AC37	AC39	0	13	CL22	AC42	0,025
33	AC14	AC29	0,003	12	AC01	CL27	0,029
32	AC08	AC09	0,004	11	CL18	AC34	0,030
31	AC30	AC31	0,004	10	CL21	CL13	0,038
30	AC28	AC40	0,004	9	CL37	CL23	0,039
29	CL39	AC12	0,004	8	CL15	AC23	0,040
28	AC36	AC38	0,005	7	CL12	CL26	0,043
27	AC02	CL41	0,005	6	CL14	CL11	0,066
26	CL40	AC07	0,005	5	CL10	CL6	0,069
25	CL33	CL34	0,006	4	CL8	CL9	0,081
24	CL32	AC25	0,008	3	CL7	CL4	0,107
23	CL36	AC33	0,008	2	CL3	AC41	0,121
22	AC15	CL31	0,008	1	CL2	CL5	0,142
21	CL29	CL35	0,008				

REFERÊNCIAS

BUSSAB, W. de O.; MIAZAKI, E.S.; ANDRADE, D.F. **Introdução à Análise de Agrupamentos**. In: 9º Simpósio Nacional de Probabilidade e Estatística, São Paulo. Associação Brasileira de Estatística, 105p.,1990.

WARD, J.H. Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**, Washington, n.58, p.236, 1963.