

PARECENÇA ENTRE ACESSOS TRADICIONAIS DE FEIJÃO CARIOCA UTILIZANDO O MÉTODO DE WARD

Jaison Pereira de **OLIVEIRA**¹
Maria José **Del PELOSO**¹
Orlando Peixoto de **MORAIS**¹
Leonardo Cunha **MELO**¹
Heloisa Torres da **SILVA**¹
Jaime Roberto **FONSECA**¹
Silvando Carlos da **SILVA**¹

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 pareências. 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 pareência. 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 pareência em 126 acessos tradicionais de feijão do grupo comercial carioca, 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 126 acessos tradicionais de feijão do grupo comercial carioca 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

¹Embrapa Arroz e Feijão, Caixa Postal 179, CEP 75375-000, Santo Antônio de Goiás, GO. E-mail: jaison@cnpaf.embrapa.br

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 $x_{kj}(i)$ tal que: $x_{kj}(i) = 1$ se $y_i = k$ ou $x_{kj}(i) = 0$. Assim o vetor y de dimensão n , é transformado no vetor x de dimensão p , formado apenas por componentes binários, e por conseqüência, 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 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 125 etapas de agrupamento (devido aos 126 acessos). Tais etapas levaram a formação de um dendrograma com arranjos distintos para diversas distâncias de agrupamento (Figura 1). Observando-se os grupos formados, notou-se que os acessos reunidos nos primeiros passos de agrupamento, cujos valores do coeficiente de aglomeração foram inferiores a 0,0007 (Tabela 1) foram semelhantes entre si, indicando possíveis redundâncias. Essas possíveis redundâncias foram observadas em 17 grupos. Nesse caso, o agrupamento 17, chama a atenção pelo número de acessos agregados. Esse grupo foi formado por 18 acessos, que possuem as mesmas características nas variáveis utilizadas para análise, ou seja, tiveram o mesmo tamanho de semente e hábito de crescimento, foram coletados na mesma região geográfica, estado da federação, classe de solo e altitude.

Os agrupamentos 6, 7, 8, e 10, respectivamente, com 4, 7, 3 e 3 acessos, formaram os grupos com maiores números de acessos, depois do grupo 17, que apresentaram redundâncias. Os demais 12 grupos possuíam dois acessos redundantes. Os outros 67 acessos tiveram pelo menos uma característica distinta. Outro ponto interessante é que os 17 grupos, com pelo menos dois acessos redundantes e formados por 59 acessos, muito provavelmente podem ser representados por apenas 17 acessos, sem perda de informações genética. Isso implica que os 126 acessos de feijão carioca podem ser resumidos à 84 acessos, sendo 17 oriundos dos

agrupamentos mais 67 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.

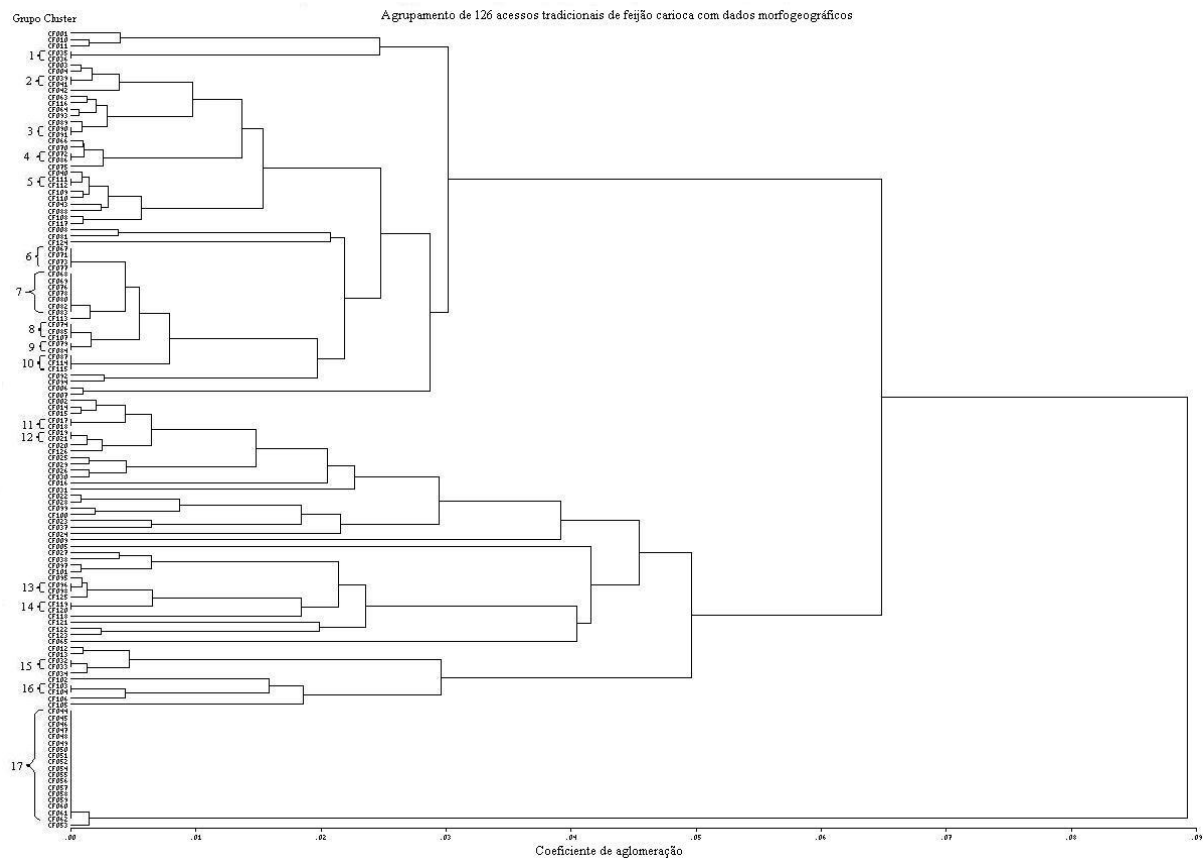


Figura 1 - Agrupamento de 126 acessos de cultivares de feijão do grupo comercial carioca, pelo método de Ward.

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 126 acessos dispersos nos 125 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.

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 50 primeiros passos do agrupamento de 126 acessos tradicionais de feijão do grupo comercial carioca, pelo método de Ward.

Estágio	Agrupamento combinado		Coeficiente de aglomeração	Estágio	Agrupamento combinado		Coeficiente de aglomeração
	Acesso/agrupamento				Acesso/agrupamento		
1	CF017	CF018	0,0000	26	CL103	CF076	0,0000
2	CF019	CF021	0,0000	27	CL101	CF077	0,0000
3	CF032	CF033	0,0000	28	CL100	CF078	0,0000
4	CF035	CF036	0,0000	29	CL98	CF080	0,0000
5	CF039	CF041	0,0000	30	CL97	CF082	0,0000
6	CF044	CF045	0,0000	31	CL96	CF083	0,0000
7	CL120	CF046	0,0000	32	CF079	CF084	0,0000
8	CL119	CF047	0,0000	33	CF074	CF085	0,0000
9	CL118	CF048	0,0000	34	CF072	CF086	0,0000
10	CL117	CF049	0,0000	35	CF090	CF091	0,0000
11	CL116	CF050	0,0000	36	CF096	CF098	0,0000
12	CL115	CF051	0,0000	37	CF103	CF104	0,0000
13	CL114	CF052	0,0000	38	CL93	CF107	0,0000
14	CL113	CF054	0,0000	39	CF111	CF112	0,0000
15	CL112	CF055	0,0000	40	CF087	CF114	0,0000
16	CL111	CF056	0,0000	41	CL86	CF115	0,0000
17	CL110	CF057	0,0000	42	CF119	CF120	0,0000
18	CL109	CF058	0,0000	43	CF064	CF093	0,0007
19	CL108	CF059	0,0000	44	CF014	CF015	0,0008
20	CL107	CF060	0,0000	45	CF003	CF004	0,0008
21	CL106	CF061	0,0000	46	CF022	CF028	0,0008
22	CL105	CF062	0,0000	47	CF097	CF101	0,0008
23	CF068	CF069	0,0000	48	CF089	CL91	0,0009
24	CF067	CF071	0,0000	49	CF095	CL90	0,0009
25	CL102	CF073	0,0000	50	CF040	CL87	0,0009

REFERÊNCIAS BIBLIOGRÁFICAS

BUSSAB, W. de O.; MIAZAKI, E.S.; ANDRADE, D.F. **Introdução à Análise de Agrupamentos**. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 9., 1990, São Paulo. São Paulo: USP-IME, 1990. 105p.

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

Área: Genética e Melhoramento