

SELEÇÃO DE DESCRITORES NA CARACTERIZAÇÃO E AVALIAÇÃO PRELIMINAR DE GERMOPLASMA DE GUANDU¹

CARLOS ANTONIO F. SANTOS², CARLOS ALBERTO V. OLIVEIRA³ e EDUARDO A. MENEZES⁴

RESUMO - Na caracterização de acessos de bancos de germoplasmas, a prática generalizada tem sido a observação de um grande número de descritores, sem avaliações da contribuição relativa de cada descritor para a discriminação fenotípica ou explicação de determinados caracteres. Para analisar os descritores importantes em guandu (*Cajanus cajan* (L.) Millsp.), foram caracterizados 54 acessos no ano de 1991 e 34 acessos no ano de 1992, na EMBRAPA-CPATSA, em Petrolina, PE, adotando-se a análise dos componentes principais para o conjunto das variáveis e a regressão múltipla, tendo como variável dependente a produção de grãos (g/planta). Na análise dos componentes principais, foram considerados como descritores discriminantes os caracteres número de sementes por vagem, peso de 100 grãos, localização da flor, e cor principal da semente. Na regressão múltipla, os caracteres mais importantes para explicação da produção de grãos/planta foram: número de vagens/planta, número de sementes/vagens, peso de 100 sementes, tamanho do hilo, forma da semente e comprimento da vagem.

Termos para indexação: *Cajanus cajan*, componentes principais, regressão múltipla, melhoramento genético.

SELECTION OF DESCRIPTORS IN CHARACTERIZATION AND PRELIMINAR EVALUATION OF PIGEONPEA GERMPLASM.

ABSTRACT- For the characterization of plant entries in germplasm collection, the evaluation of a big number of descriptors without evaluating the relative contribution of each one has been a generalized practice. In order to analyze the discriminant descriptors on pigeonpea (*Cajanus cajan* (L.) Millsp.), 54 entries were characterized in 1991 and 34 in 1992 at EMBRAPA-CPATSA, Petrolina, PE, Brazil, using the analysis of principal components for the set of variables and the multiple regression, with grain yield (g/plant) as the dependent variable. In the analysis of principal components, the variables number of seeds/pod, 100-seed weight, flower location and seed main color were considered. In the multiple regression analysis, the most important variables for explaining grain yield per plant were: number of pods/plant, number of seed/pod, 100-seed weight, hilum size, seed shape and pod length.

Index terms: *Cajanus cajan*, principal components, multiple regression, genetic breeding.

INTRODUÇÃO

Na caracterização de acessos de germoplasma, a prática generalizada é a avaliação de um grande número de descritores, tanto qualitativos como quantitativos. Avaliações da importância relativa de cada descritor para discriminação fenotípica ou ex-

plicação de caracteres de importância agrônômica, como a produção de grãos, não têm sido enfatizadas.

Caracteres considerados como dispensáveis no estudo da divergência genética são os relativamente invariantes entre as cultivares avaliadas e/ou explicadas por outros caracteres. O descarte destes caracteres propicia redução de mão-de-obra, tempo e custo (Cruz & Regazzi, 1994), contribuindo ainda para a mensuração mais acurada das variáveis importantes para a caracterização.

Para o IBPGR (1981), na caracterização de acessos, são realizadas avaliações de caracteres de alta herdabilidade, facilmente observados e que se ex-

¹ Aceito para publicação em 19 de junho de 1995.

² Eng. Agr., M.Sc. EMBRAPA-CPATSA Caixa Postal 23 CEP 56300-000 Petrolina-PE.

³ Estatístico, EMBRAPA-CPATSA.

⁴ Eng. Agr., Ph.D. EMBRAPA-CPATSA.

pressam em todos os ambientes, enquanto na avaliação preliminar são realizadas também mensurações de alguns caracteres de importância agrônômica, como a produção de grãos, por exemplo. Estes dois procedimentos podem ser realizados concomitantemente, numa só etapa de avaliação, se o número de acessos avaliados não for muito grande.

Técnicas de análises multivariadas, como regressão (Beale et al., 1967), análise discriminante (Mardia et al., 1979), componentes principais e variáveis canônicas (Cruz & Regazzi, 1994), são sugeridas para a seleção ou descarte de variáveis.

Comparações entre a consistência de algumas destas técnicas não são ainda conclusivas (Jolliffe, 1973), necessitando de estudos mais aprofundados.

Neste trabalho, procura-se selecionar os descritores importantes para a caracterização e avaliação preliminar de germoplasma de guandu (*Cajanus cajan* (L.) Millsp), por meio da técnica dos componentes principais e da regressão múltipla.

MATERIAL E MÉTODOS

Foram caracterizados, na EMBRAPA-CPATSA, em Petrolina, PE, 88 acessos de guandu, sendo 54 no ano de 1991 e 34 no ano de 1992, segundo método adaptado do IBPGR (1981). As avaliações foram efetuadas em parcelas únicas, sem repetição, considerando-se os mesmos 20 descritores nas duas avaliações. Os caracteres mensurados por intermédio de escalas tiveram os seus valores somados a um valor fixo e transformados em raiz quadrada.

Foram avaliadas as seguintes técnicas multivariadas para seleção dos descritores discriminantes ou explicativos da produção de grãos em guandu:

1) Análise dos componentes principais: consiste em transformar um conjunto original de variáveis em outro conjunto de dimensão equivalente, ambos independentes entre si, e retendo o máximo de informação em ordem de estimação. Cruz & Regazzi (1994) descrevem esta técnica de forma pormenorizada.

Consideraram-se, os menores autovalores, para descartar a variável com maior valor absoluto no respectivo autovetor, como descrito por Mardia et al. (1979). O processo continua até o autovalor de 0,7, como recomendado por Jolliffe (1972, 1973). Foram selecionados como descritores que mais discriminaram os caracteres não

descartados em nenhuma das duas avaliações (Daher et al., 1993). Estas análises foram executadas pelo programa computacional "Genes".

2) Análise de regressão múltipla: considerou-se como variável dependente a produção de grãos (g/planta), e os demais descritores, como variáveis independentes.

Para determinar a ordem de importância das variáveis no modelo explicativo da produção, foram testados 4 métodos de seleção de variáveis, quais sejam: "backward", "forward selection", "stepwise" e o "R² máximo". Todos esses métodos foram consistentes e apresentaram resultados semelhantes.

Neste trabalho, apresenta-se apenas o "forward selection" (Draper & Smith, 1966). Nesse método, a ordem de inserção das variáveis no modelo é determinada pelos coeficientes de correlações parciais. O procedimento básico é: 1) seleciona-se o Xi mais correlacionado com Y e determinam-se a equação de regressão F (Xi); 2) determinam-se os coeficientes parciais de correlação Xj (j ≠ i) e Y. O Xj com maior correlação parcial com o Y é selecionado, ajustando-se uma segunda equação de regressão Y=f(X1, X2). Esse critério continua até a detecção das variáveis independentes não-significativas pelo teste "F". Estas análises foram executadas pelo software "Statistical Analyses Systems"(SAS).

RESULTADOS E DISCUSSÃO

Na Tabela 1, são apresentados os autovalores e variáveis com maior coeficiente nos respectivos autovalores, obtidos por intermédio da técnica dos componentes principais. Observa-se que os caracteres que mais discriminaram os acessos de guandu foram NSV (número de sementes/vagem), PCG (peso de 100 grãos), LOC (localização da flor) e CPS (cor principal da semente), pois não foram descartados em nenhuma das duas avaliações.

Os caracteres RAM (ramificação), CVA (comprimento da vagem), CPV (cor da vagem), PRO (produção de grãos/planta), e ALP (altura da planta) foram descartados nas duas avaliações (Tabela 1). Pode-se considerá-los como os de menor importância na discriminação dos acessos estudados. Os demais caracteres podem ser considerados como de importância intermediária, pois foram descartados em apenas uma avaliação.

Os caracteres ALP (altura da planta) e AFL (abscisão foliar), apesar de apresentarem maior peso

TABELA 1. Autovalores e variáveis de maior coeficiente nos respectivos autovetores estimados através da técnica de componentes principais, em dados de 54 e 34 acessos de guandu caracterizados^{1/} nos anos de 1991 e 1992, respectivamente.

Auto-valores		Variáveis	
1991	1992	1991	1992
4.5069	3.7991	ALP	AFL
2.6181	2.7323	NVP	THI
2.3344	2.2879	CPV	PCG
1.8668	1.7661	CPS	ALP
1.5039	1.6543	LOC	CFL
1.2757	1.4507	AFL	NVP
0.9533	1.3674	ALR	LOC
0.7901	0.8935	ALR	CSE
0.7765	0.8272	CTL	RAM
0.6750	0.6463	RAM	NVP
0.5427	0.5144	FSE	ALR
0.4586	0.5014	THI	CFL
0.4071	0.4654	DEV	CVA
0.3303	0.2789	CPV	CSE
0.2905	0.2439	RAM	RAM
0.2319	0.1935	CVA	CPV
0.2095	0.1372	CTL	DPC
0.1432	0.1108	DPF	ALP
0.0687	0.0827	ALP	AFL
0.0163	0.0469	PRO	PRO

^{1/} RAM=ramificação; CTL=cor do talo; CFL=cor da flor; LOC=localização da flor; DEV=distância entre vagens; ALR=altura do primeiro ramo; ALP=altura da planta; AFL=abscisão foliar; DPF=dias para floração; CPV=côr da vagem; CVA=comprimento da vagem; NSV=número de sementes/vagem; CSE=distribuição da cor na semente; CPS=cor principal da semente; THI=tamanho do hilo; FSE=forma da semente; PCG=peso de 100 grãos; DPC=dias para a colheita; PRO=produção de grãos e NVP=número de vagens/planta.

nos maiores autovalores das avaliações de 1991 e 1992, respectivamente, foram descartados (Tabela 1). Esse fato demonstra certa inconsistência da técnica empregada.

Na Tabela 2, são apresentadas as variáveis selecionadas e os coeficientes de determinação acumulados (R^2), na análise de regressão múltipla para explicação da produção de grãos de guandu. Observa-se que os componentes primários NVP (número de sementes/planta), NSV e PCG (Santos et al., 1994), apresentaram R^2 de 94,00% e 97,96% nas análises de 1991 e 1992, respectivamente. Esse fato destaca a importância desses componentes para explicação da produção, apesar da inclusão do caráter THI (tamanho do hilo) como a terceira variável do modelo em 1991.

Quando o caráter NVP é considerado isoladamente (Tabela 2), observa-se que ele foi responsável pela explicação de mais de 74% da produção de grãos de guandu nas duas avaliações.

Além dos componentes primários, outras variáveis como THI, FSE (forma da semente) e CVA foram importantes para o modelo explicativo da produção, nas duas avaliações efetuadas (Tabela 2), podendo ser consideradas no processo de avaliação preliminar de acessos de guandu.

Considerando o conjunto das análises (Tabelas 1 e 2), observa-se que os caracteres número de sementes/vagem, peso de 100 grãos, localização da flor e cor principal da semente foram importantes para discriminação fenotípica, enquanto os caracteres número de vagens/planta, peso de 100 grãos, número de sementes/vagem, tamanho do hilo, forma da semente e comprimento da vagem foram importantes para explicação da produção de grãos do guandu. Dos 20 descritores inicialmente avaliados, apenas sete foram considerados como discriminantes dos acessos ou explicativos da produção de grãos em guandu.

Os caracteres identificados na análise dos componentes primários e os selecionados na análise de regressão múltipla podem ser priorizados como descritores nos trabalhos de caracterização e avaliação preliminar de germoplasma de guandu.

TABELA 2. Variáveis selecionadas e coeficiente de determinação acumulado (R^2) na análise de regressão múltipla, considerando a variável produção de grãos (g/planta) como a variável dependente, em dados de 54 e 34 acessos de guandu, caracterizados para 19 descritores^{1/}, nos anos de 1991 e 1992, respectivamente. Petrolina, PE.

Variáveis selecionadas		R^2	
1991	1992	1991	1992
NVP	NVP	0.7477	0.7596
NVP,PCG	NVP,NSV	0.9118	0.8602
NVP, PCG,THI	NVP,NSV,PCG	0.9213	0.9796
NVP,PCG,THI,NSV	NVP,NSV,PCG,CFL	0.9400	0.9822
NVP,PCG,THI,NSV,CSE	NVP,NSV,PCG,CFL,DPC	0.9491	0.9842
NVP,PCG,THI,NSV,CSE,CTL	NVP,NSV,PCG,CFL,DPC,DPF	0.9524	0.9859
NVP,PCG,THI,NSV,CSE,CTL,FSE	NVP,NSV,PCG,CFL,DPC,DPF,CVA	0.9555	0.9867
NVP,PCG,THI,NSV,CSE,CTL,FSE, CVA	NVP,NSV,PCG,CFL,DPC,DPF,CVA, FSE	0.9572	0.9876
NVP,PCG,THI,NSV,CSE,CTL,FSE, CVA,CPS	NVP,NSV,PCG,CFL,DPC,DPF,CVA, FSE,THI	0.9588	0.9882
NVP,PCG,THI,NSV,CSE,CTL,FSE, CVA,CPS,CPV	NVP,NSV,PCG,CFL,DPC,DPF,CVA, FSE,THI,AFL	0.9601	0.9894
NVP,PCG,THI,NSV,CSE,CTL,FSE, CVA,CPS,CPV,ALP	NVP,NSV,PCG,CFL,DPC,DPF,CVA, FSE,THI,AFL,LOC	0.9612	0.9903
NVP,PCG,THI,NSV,CSE,CTL,FSE, CVA,CPS,CPV,ALP,LOC	NVP,NSV,PCG,CFL,DPC,DPF,CVA, FSE,THI,AFL,LOC,RAM	0.9617	0.9906
	NVP,NSV,PCG,CFL,DPC,DPF,CVA, FSE,THI,AFL,LOC,RAM,CSE		0.9908

^{1/} RAM=ramificação; CTL=cor do talo; CFL=cor da flor; LOC=localização da flor; DEV=distância entre vagens; ALR=altura do primeiro ramo; ALP=altura da planta; AFL=abscisão foliar; DPF=dias para floração; CPV=cor da vagem; CVA=comprimento da vagem; NSV=número de sementes/vagem; CSE=distribuição da cor na semente; CPS=cor principal da semente; THI=tamanho do hilo; FSE=forma da semente; PCG=peso de 100 grãos; DPC=dias para a colheita e NVP=número de vagens/planta.

CONCLUSÕES

1. No estudo dos componentes principais, os caracteres que mais discriminaram os acessos de guandu foram: número de sementes/vagem, peso de 100 grãos, localização da flor e cor principal da semente.

2. No estudo da regressão múltipla, os caracteres de maior importância para explicação da produção de grãos em guandu foram: número de vagens/planta, número de sementes/vagem, peso de 100 sementes, tamanho do hilo, forma da semente e compri-

mento da vagem.

3. A avaliação concomitante dos caracteres importantes para a discriminação fenotípica e a explicação de grãos em guandu possibilitou reduzir a lista inicial de vinte descritores para apenas sete.

4. Os caracteres número de sementes/vagem, peso de 100 grãos, localização da flor, cor principal da semente, número de vagens/planta, tamanho do hilo, forma da semente e comprimento da vagem podem ser priorizados nos trabalhos de caracterização e avaliação preliminar de germoplasma de guandu.

REFERÊNCIAS

- BEALE, E.M.; KENDALL, M.G.; MANN, D.W. The discarding of variables in multivariate analysis. **Biometrika**, v.54, n.3/4, p.357-366, 1967.
- CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, Imp. Univ., 1994.
- DAHER, R.F.; MORAES, C.F.; PEREIRA, A.V.; XAVIER, D.F. Seleção de caracteres discriminantes em germoplasma de capim elefante (*Pennisetum purpureum* Schum.). **Revista Brasileira de Genética**, v.16, n.3, p.349, 1993. Suplemento.
- DRAPER, N.R.; SMITH, H. **Applied regression analysis**. New York: J. Wiley, 1966. 407p.
- IBPGR. **Descriptors for Pigeonpea**. Rome, Italy, 1981. 15p.
- JOLLIFFE, I.T. Discarding variables in a principal component analysis. I. Artificial data. **Applied Statistics**, v.21, p.160-173, 1972.
- JOLLIFFE, I.T. Discarding variables in a principal component analysis. II. real data. **Applied Statistics**, v.22, p.21-31. 1973.
- MARDIA, R.F.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. [S.l.]: Academic Press, 1979. 521p.
- SANTOS, C.A.F.; MENEZES, E.A.; PAINI, C.D.; CRUZ, C.D. Coeficiente de trilha no estudo dos componentes primários e secundários na produção de grãos do Guandu (*Cajanus cajan* (L.) Millsp). **Revista Ceres**, v.41, n.235, p.299-305, 1994.