



AGRUPAMENTO NÃO HIERÁRQUICO DE ACESSOS DE GERMOPLASMA DE MANDIOCA (*Manihot esculenta* Crantz) COM BASE EM CARACTERÍSTICAS QUANTITATIVAS

Eder Jorge de Oliveira¹, Fabiana Aud², Cinara Fernanda Garcia Morales², Vanderlei da Silva Santos¹, Fabiana Kotwiski³, Vanderlei Costa³

¹Pesquisador da *Embrapa Mandioca e Fruticultura*, Caixa Postal 007, 44380-000, Cruz das Almas, BA. E-mail: eder.oliveira@embrapa.br, vanderlei.silva-santos@embrapa.br

²Analista da *Embrapa Mandioca e Fruticultura*, Caixa Postal 007, 44380-000, Cruz das Almas, BA. E-mail: fabiana.aud@embrapa.br, cinara.morales@embrapa.br

³Analista da *Bahiamido*, 44571 970, Laje, BA. E-mail: fabiana@bahiamido.com.br, vanderlei@bahiamido.com.br

Introdução

Os recursos genéticos de mandioca (*Manihot esculenta* Crantz) são utilizados como fonte de variação genética para incorporação de genes de elevado interesse agrônomico em novas variedades de mandioca. Muitas características importantes ainda nem foram descobertas neste germoplasma, mas de maneira geral a busca por novos alelos para características tradicionalmente importantes no sistema de produção da mandioca como aquelas relacionadas ao aumento de produtividade, teor e qualidade de amido, resistência à seca, pragas e doenças ainda continua sendo uma linha de pesquisa extremamente importante como uma etapa que antecede o seu uso nos programas de melhoramento.

O conhecimento da variação fenotípica do germoplasma de mandioca permite conhecer a real variabilidade genética, e subsidiar a seleção de parentais geneticamente divergentes, que poderão ser utilizados em intercruzamentos para se obter alto efeito heterótico e aumentar assim a probabilidade de recuperação de segregantes superiores em gerações avançadas. Embora diversos métodos estejam disponíveis para analisar a variação fenotípica, alguns métodos não hierárquicos como o *k-means* tem se destacado no agrupamento de genótipos especialmente quando há um grande conjunto de dados. Assim, o objetivo deste trabalho foi definir agrupamentos homogêneos de acessos de germoplasma de mandioca com base em características produtivas e de qualidade da raíz.

Material e Métodos

O ensaio de campo foi instalado em setembro de 2011 na Cooperativa de Produtores de Amido (Coopamido) no município de Laje (BA). O delineamento experimental foi de blocos aumentados com 629 acessos de germoplasma de mandioca como tratamentos não

comuns e 11 testemunhas como tratamentos comuns, distribuídos em 10 blocos com parcela de 10 plantas. Os acessos são oriundos do Banco Ativo de Germoplasma de Mandioca (BAG-Mandioca) da Embrapa Mandioca e Fruticultura, localizada em Cruz das Almas (BA). O espaçamento utilizado foi de 0,90m entre linhas e 0,80m entre plantas e os tratamentos culturais foram realizados de acordo com recomendações da cultura (SOUZA et al., 2006). A colheita foi realizada aos 20 meses de plantio.

Os dados avaliados em t.ha⁻¹ foram peso de raízes comerciais (PRC) e não comerciais (PRNC); peso da parte aérea (PPA); produtividade de amido (PROD-AMD - considerando o teor de amido e a produtividade total de raízes). Também foram avaliados o índice de colheita (IC), que caracteriza a relação entre produção de raízes e a biomassa aérea da mandioca, mensurado em %; altura de plantas (AP), mensurada em m; renda, mensurada pela pesagem em balança hidrostática de 5 kg de raízes frescas submersas em água; teor de matéria seca (MS), em %; teor de amido (AMD), em %; severidade da bacteriose por meio de escala de notas (0 a 4), antracnose (0 a 4) e avaliação da capacidade de recuperação da planta após a infestação por bacteriose e antracnose (0 a 2).

Os valores genéticos obtidos por meio do melhor preditor linear não viesado (*Best Linear Unbiased Predictor* – BLUP) foram utilizados para o agrupamento dos acessos utilizando o *k-means*. Neste método é realizado um agrupamento não hierárquico de forma a minimizar a distância dos elementos a um conjunto de *k* centros dado por $\chi = \{x_1, x_2, \dots, x_k\}$ de forma iterativa. A distância entre um ponto p_i e um conjunto de clusters, dada por $d(p_i, \chi)$, é definida como sendo a distância do ponto ao centro mais próximo dele. A função a ser

minimizada então, é dada por: $d(P, \chi) = \frac{1}{n} \sum_{i=1}^n d(p_i, \chi)^2$.

Resultados e Discussão

O número de grupos foi definido em função da estabilização da menor soma de quadrados dentro. Assim, de acordo com Figura 1, foram definidos 17 grupos de acessos de mandioca para representar a diversidade genotípica observada. A escolha deste número de grupo revela alta variabilidade genotípica neste conjunto de germoplasma analisado, cujos valores médios de cada grupo são apresentados na Tabela 1. O número de acessos por grupo variou de 7 (Grupo 15) a 69 (Grupo 9), com média de 37,6.

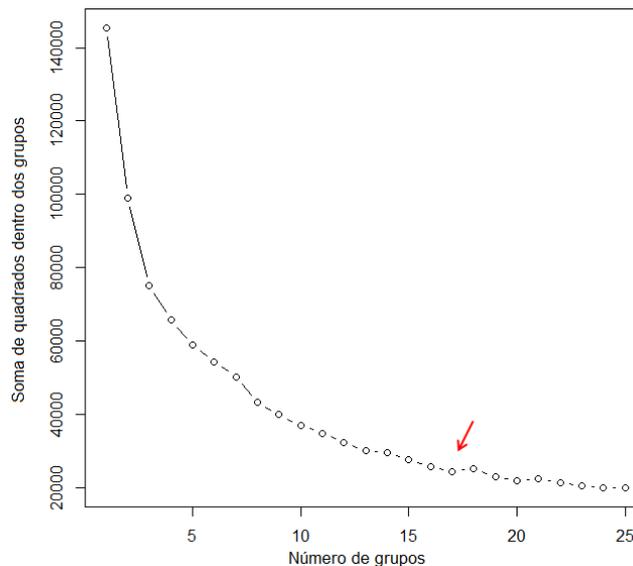


Figura 1. Distribuição da soma de quadrados dentro em função do número de grupos.

Tabela 1. Distribuição acessos de mandioca, com base nos valores genotípicos obtidos pelo BLUP, por meio da análise de 12 características agrônômicas.

Grupo	Nº indiv	Características ^a												SQd ^b
		1	2	3	4	5	6	7	8	9	10	11	12	
1	27	65,6	9,1	37,2	66,0	2,4	498,4	32,8	28,1	20,4	1,4	3,1	1,3	1538,8
2	37	40,6	9,2	27,2	65,4	2,3	317,3	27,0	22,4	12,3	1,2	3,1	1,3	1589,7
3	14	41,4	9,8	43,6	55,4	2,4	488,3	32,4	27,7	14,3	1,3	3,1	1,3	1397,7
4	25	51,8	9,5	39,9	61,4	2,4	503,2	32,9	28,3	17,0	1,4	3,1	1,3	1040,4
5	59	37,4	9,0	23,8	66,8	2,2	404,3	29,9	25,2	12,1	1,3	3,1	1,3	1742,5
6	47	44,6	9,4	31,4	63,4	2,4	551,0	34,4	29,8	15,2	1,3	3,1	1,3	1367,4
7	9	74,2	9,4	44,8	65,5	2,6	532,4	33,9	29,2	23,7	1,5	3,0	1,3	661,6
8	58	36,7	9,0	27,3	61,6	2,3	521,5	33,5	28,9	12,4	1,3	3,1	1,3	2338,3
9	69	41,3	9,3	29,7	62,9	2,3	429,2	30,5	25,9	13,3	1,3	3,1	1,3	1660,1
10	34	34,4	8,9	28,2	57,1	2,3	421,0	31,1	26,5	11,4	1,3	3,1	1,3	1878,4
11	56	48,0	9,3	30,5	65,6	2,3	467,8	31,7	27,0	15,3	1,3	3,1	1,3	1197,0
12	30	56,5	9,3	29,8	68,7	2,3	517,1	33,4	28,7	18,3	1,3	3,1	1,3	1132,3
13	34	48,2	9,4	29,3	67,1	2,3	376,1	28,5	23,9	14,5	1,2	3,1	1,3	1230,9
14	25	35,5	8,8	28,9	57,4	2,3	297,9	27,5	22,8	11,1	1,2	3,1	1,3	1281,8
15	7	58,3	9,1	36,1	64,6	2,4	314,3	26,6	21,9	15,5	1,0	3,1	1,3	215,1
16	43	57,6	9,1	31,7	67,3	2,4	419,7	30,1	25,5	17,0	1,3	3,1	1,3	1486,5
17	65	41,8	9,3	25,7	67,7	2,2	489,4	32,4	27,8	13,9	1,2	3,1	1,3	1569,3

^a: os números de 1 a 12 referem-se a PRC, PRNC, PPA, IC, AP, RENDA, MS, AMD, PROD.AMD, Bacteriose, Antracnose e RECPL, respectivamente, conforme descrito em Material e Métodos; ^b: SQd é a soma de quadrados dentro dos grupos.

O método *k-means* particiona os dados em *k* agrupamentos mutuamente exclusivos, não havendo a construção de dendogramas para descrever o agrupamento dos dados. Contudo, é possível representar a distribuição dos indivíduos e das características avaliados com base na análise de componentes principais (Figura 2).

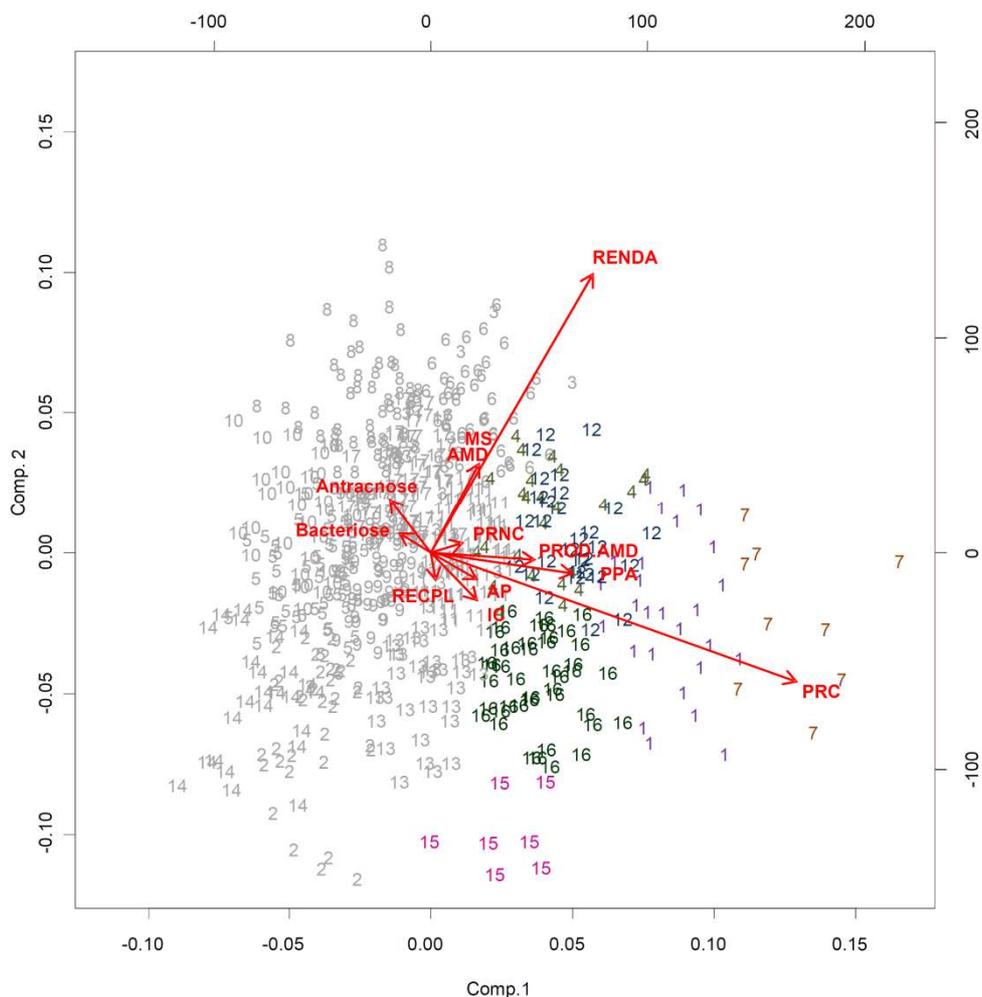


Figura 2. Distribuição da soma de quadrados dentro em função do número de grupos.

De modo geral, observa-se que os acessos pertencentes aos Grupos 1, 4, 7, 12, 15 e 16 apresentam características agrônômicas para uso direto no sistema de produção, ou como parentais em blocos de cruzamento. Por exemplo, nestes grupos a produtividade média de raízes comerciais e de amido foi de 60,7 t.ha⁻¹ e 18,6 t.ha⁻¹, que representa cerca de 26,7% e 22,3%, respectivamente, superior à médias dos demais agrupamentos (Tabela 1). Com relação à severidade das doenças, não foram identificados grandes diferenças na média dos

grupos selecionados em relação aos demais, à exceção do Grupo 15, na qual o foco principal da escolha deveu-se à menor severidade da bacteriose, muito embora sejam acessos com menor renda de raízes e por consequência menor matéria seca e amido.

Observou-se a partição dos grupos com padrões bastante próximos entre si e ao mesmo tempo distantes dos padrões dos outros agrupamentos. Isto pode ser demonstrado pelos reduzidos valores da soma de quadrados dentro dos grupos, que variou de 215,1 (Grupo 15) a 2.338,3 (Grupo 8), com média de 1.372,2. Isto representa em média apenas 1,2% da soma de quadrados total (109.754,7). O *k-means* é um algoritmo que minimiza a soma das distâncias de cada padrão ao centróide de cada agrupamento, sobre todos os agrupamentos, culminando com a formação de grupos bastante consistentes, como os observados na análise do germoplasma de mandioca.

Conclusões

Agrupamentos não hierárquicos como o *k-means* constituem um método eficiente para a formação de grupos com a menor variação genotípica dentro, especialmente quando se tem uma grande quantidade de dados, como em bancos de germoplasma de mandioca.

Agradecimentos

Os autores agradecem ao CNPq e à Aliança Cooperativa do Amido pelo apoio financeiro e suporte necessário para realização desta pesquisa.

Referências

SOUZA, L.S., FARIAS, R.N. **Aspectos socioeconômicos e agronômicos da mandioca.** Cruz das Almas – BA. Embrapa Mandioca e Fruticultura Tropical, 817 p., 2006.