

USO DE MARCADORES RAPD E SSR NA DETECÇÃO DE ACESSOS DUPLICADOS DE MANDIOCA (*Manihot esculenta* Crantz)

Ferreira, CF¹, Souza, RMS.², Silva, PH.³, Ledo, CAS¹

¹Pesquisadora da Embrapa Mandioca e Fruticultura Tropical, S/N, CP 007, 44380-000 Cruz das Almas-BA. Tel: (75)-3621-8061, Fax: (75)-3634-8097, Laboratório de Virologia e Biologia Molecular. E-mail: claudiaf@cnpmf.embrapa.br/ledo@cnpmf.embrapa.br.

²Universidade Estadual do Maranhão, UEMA - Cidade Universitária Paulo VI – Caixa Postal 09–São Luís-MA. Tel: (98)-3245-5461, Fax: (98)-3245-5882, Curso de Biologia.

⁵Universidade Federal de Lavras, UFLA, CP 3037, Lavras, MG, telefone: (35) 3829-1122. E-mail: p p h s i l v a @ h o t m a i l . c o m

INTRODUÇÃO

A mandioca (*Manihot esculenta* Crantz) apresenta elevada diversidade genética concentrada principalmente na América Latina e Caribe. Aproximadamente 8.500 acessos são mantidos em bancos de germoplasma em várias coleções espalhadas mundialmente, dos quais 7.500 encontram-se na América do Sul. No Brasil, cerca de 4.132 acessos já foram coletados e estão sendo mantidos em bancos de germoplasma em vários locais do País.

A Embrapa Mandioca e Fruticultura, sediada em Cruz das Almas – BA, possui o maior Banco de Germoplasma de Mandioca (BAG)-Mandioca do Brasil, com 1300 acessos em campo, numa área total de 3 ha. As principais atividades desenvolvidas no banco compreendem a coleta, conservação, caracterização, avaliação, documentação e intercâmbio, servindo de apoio para os principais programas de melhoramento da cultura no país.

A mandioca é uma espécie altamente heterozigótica, cuja forma principal de propagação é vegetativa. A identificação de duplicatas é considerada um passo importante no desenvolvimento de uma coleção nuclear capaz de representar com confiabilidade, a variabilidade genética disponível da espécie, otimizando a melhor condução dos trabalhos pelos melhoristas e tornando o germoplasma mais acessível a todos os seus usuários.

Embora o banco de germoplasma de mandioca da Embrapa Mandioca e Fruticultura seja o maior do Brasil, o mesmo vem enfrentando dificuldades devido ao elevado custo de manutenção em decorrência da limitação de mão-de-obra e ataque de pragas e doenças. Outro agravante é a ocorrência de acessos duplicados, que até o presente vem sendo avaliada apenas por características morfológicas, podendo gerar dúvidas. O custo para se manter um acesso de mandioca em nível de campo, por ano, está em torno de US\$ 17.66 e *in vitro*, US\$ 24.70 (Epperson et al. 1996).

Portanto, o principal objetivo do trabalho foi identificar, em caráter preliminar, duplicatas entre sessenta e um acessos de mandioca pertencentes ao BAG-Mandioca, já caracterizados morfológicamente, utilizando-se marcadores do tipo RAPD e SSR de forma a viabilizar a manutenção dos acessos representativos do banco e com isso viabilizar as tarefas de armazenamento e intercâmbio de materiais entre instituições nacionais e internacionais.

MATERIAL E MÉTODOS

Material genético

Sessenta e um acessos de mandioca provenientes do BAG-Mandioca da Embrapa Mandioca e Fruticultura foram utilizados na identificação de acessos duplicados. Esses acessos já haviam sido separados em grupos de acordo com descritores morfológicos.

Extração e amplificação de DNA genômico

A extração de DNA foi feita a partir de folhas jovens de 61 acessos de mandioca anteriormente separados em grupos de acordo com as características morfológicas seguindo a metodologia descrita por Doyle & Doyle (1990). As amplificações seguiram a metodologia proposta por Williams et al. (1990). Os fragmentos de RAPD foram avaliados em gel de agarose a 1,5% e os fragmentos amplificados via SSR foram separados em gel desnaturante de poliacrilamida a 5% e submetidos à coloração com nitrato de prata segundo o método descrito por Creste (2001).

Avaliação dos dados moleculares e eliminação de duplicatas

Apenas as bandas polimórficas mais nítidas foram utilizadas no estudo. As bandas foram caracterizadas como presença (1) e ausência (0). Os dados foram compilados para uso no software GENES (Cruz, 2003) para cálculo da dissimilaridade genética utilizando o índice de Jaccard. Procedeu-se então ao cálculo da matriz híbrida, ou seja, utilizando-se os dados morfológicos e de marcadores de DNA (RAPD e SSR) simultaneamente. Esta matriz foi obtida pelo programa R, utilizando-se o algoritmo de GOWER (Gower, 1971). As distâncias genéticas foram utilizadas para realizar o agrupamento dos genótipos pelo método UPGMA (*unweighted pair-group method averages*) por meio do programa Statistica 7 e o desvio padrão calculado de acordo com a metodologia proposta por Fu et al. (2006). Os acessos considerados mais similares de acordo com os dados da matriz de dissimilaridade foram separados e as matrizes de dissimilaridades recalculadas após cada eliminação de um dos acessos, tendo sempre o cuidado de checar o índice de diversidade genética (*gene diversity*) utilizando-se o software POWERMARKER (Liu & Muse, 2005) para não prejudicar a variabilidade disponível para os melhoristas. A matriz de distância com base nos dados das variáveis multicategóricas (Cole-Rodgers, 1997) foi calculada utilizando-se o programa GENES (Cruz, 2003). O dendrograma calculado com base nesta matriz utilizando-se o software MEGA – 4 (Tamura et al. 2007),

RESULTADOS E DISCUSSÃO

Atualmente o Banco de Germoplasma de Mandioca da Embrapa Mandioca e Fruticultura conta com 1300 acessos, representando o maior banco de mandioca do Brasil. Da coleção de trabalho,

aproximadamente 40 acessos são os mais utilizados dentro do programa de melhoramento genético da cultura, demonstrando que há muito material ainda a ser avaliado para poder fazer parte desta coleção e contribuir para a ampliação da base genética da cultura.

A caracterização molecular para fins de eliminação de acessos duplicados em bancos de germoplasma é uma atividade que vem crescendo a cada dia; principalmente em virtude dos benefícios trazidos em termos de custo e manutenção dos mesmos (Virk et al, 1995; Fu et al, 2003; Van Hintum et al., 2003; Chiorato et al., 2006; Fu, 2006). O principal desafio é conseguir identificar os acessos duplicados, redundantes, eliminá-los e manter a variabilidade genética dentro do banco de germoplasma.

A dissimilaridade entre acessos por meio de marcadores de DNA pode ser medida a partir das médias das dissimilaridades (Chiorato et al., 2006, Hintum, 2003).

Foram caracterizados molecularmente 61 acessos do BAG-mandioca da Embrapa Mandioca e Fruticultura. Esses acessos foram escolhidos por estarem devidamente já caracterizados morfológicamente de acordo com os principais descritores para a cultura e, portanto sendo considerados possíveis duplicatas dentro de seus grupos baseando-se apenas nos dados morfológicos.

Primeiramente procedeu-se à análise mais detalhada dos dados morfológicos para verificar a separação feita dos 61 acessos em 20 grupos. demonstrou que os acessos continuaram agrupados nos seus respectivos grupos; demonstrando boa precisão na avaliação morfológica dos acessos.

Avaliação molecular utilizando-se marcadores RAPD e SSRs

Os 61 acessos foram então caracterizados molecularmente utilizando-se marcadores do tipo RAPD (*Random Amplified Polymorphic DNA*) e SSRs (*Simple Sequence Repeats*). Oito primers de RAPD da série Operon Technologies (Alameda, CA, USA) geraram 85 bandas polimórficas. Os quatro primers SSR utilizados foram disponibilizados pelo CIAT (Centro Internacional de Agricultura Tropical-Cali-Colombia). Estes quatro primers geraram 18 bandas polimórficas, perfazendo um total de 102 bandas polimórficas analisadas.

O número de bandas polimórficas por primer de RAPD variou de 5 bandas para o primer OPQ-8 a 17 bandas para o primer OPQ-18. Em relação aos marcadores SSR, o número de alelos variou de 3 para o primer SSRY-34 a 5 para o primer SSRY-63.

A matriz de distância genética considerando os dados dos marcadores RAPD e SSR foi obtida utilizando-se o índice de dissimilaridade de Jaccard (Jaccard, 1901) e o programa GENES (Cruz, 2003), sendo que os dados foram computados como ausência (0) e presença (1).

Com base nos dados moleculares, os acessos pertencentes ao grupo 1 continuaram próximos, bem como os dos grupos 14, 12, 16 e 17. Entretanto, a maioria dos acessos mostrou uma associação diferente da feita apenas com as características morfológicas, demonstrando a importância da complementação da avaliação morfológica com a molecular.

Procedeu-se então ao cálculo da matriz híbrida, ou seja, utilizando-se os dados morfológicos e de marcadores de DNA (RAPD e SSR). Esta matriz foi obtida pelo programa R, utilizando-se o algoritmo de GOWER (Gower, 1971).

A validação dos agrupamentos foi determinada pelo coeficiente de correlação cofenético (Soka e Rohlf, 1962), calculada pelo programa R e foi $r = 0,73$, valor acima de 0,5, indicando boa precisão e construiu-se o dendrograma com os dados híbridos.

Essa matriz híbrida forneceu os dados para os cálculos das dissimilaridades médias entre os acessos. A dissimilaridade média variou de 0,2644 a 0,3512, com valor de média geral de 0,3028 e desvio padrão de 0,0151. Isto significa que os acessos com valores de dissimilaridade média abaixo de 0,2877 ($0,3028 - 0,0151 = 0,2877$) podem ser considerados redundantes. Portanto, os acessos de mandioca que podem ser retirados do banco por serem redundantes encontram-se na Tabela 2. Estes 14 acessos representam aproximadamente 20% do total de acessos analisados.

Embora os acessos do grupo I, por exemplo, que se permaneceram unidos em todas as análises; tanto para a análise com dados morfológicos, quanto para moleculares e a em conjunto, a diferença de dissimilaridade média dentro do grupo variou de 0,2913 a 0,3030, não sendo suficiente para eliminar quaisquer componente do grupo. Isto porque a dissimilaridade média em relação a todos os acessos foi de 0,3028 com um desvio padrão de 0,0151 e nenhum acesso deste grupo apresentou dissimilaridade média em relação aos outros acessos menor que 0,2877. Isto demonstra a importância de se analisar os dados da matriz de dissimilaridade e não apenas a disposição dos acessos no dendrograma, que certamente levaria a crer que pelo menos 5 a 6 acessos do grupo 1 seriam considerados como duplicatas e eliminados erroneamente (dados não apresentados).

Entretanto, para confirmar que a diversidade genética do banco não foi alterada com a retirada dos acessos redundantes (BGM 1158, BGM 0367, BGM 0070, BGM 1156, BGM 0210, BGM 0195, BGM 1136, BGM 1294, BGM 0822, BGM 0279, BGM 0286, BGM 0878, BGM 0962, BGM 1306), calculou-se a *gene diversity* (H) antes e depois da retirada dos mesmos. Esta diversidade genética, *gene diversity* (H), é usada para se medir a diversidade genética em acessos, conjunto de acessos ou uma coleção de acessos. Se p é a frequência da banda, a diversidade para aquele marcador, $H = 1 - (p^2 + (1-p)^2) = 2p(1-p)$ (Nei, 1987).

Para um cálculo mais preciso, utilizou-se os dados dos 18 alelos dos SSRs. A medida da diversidade genética (H) calculada para os 61 acessos utilizando-se o software POWERMARKER (Liu e Muse, 2005) foi de 0,5937 e após a retirada dos acessos redundantes (BGM 1158, BGM 0367, BGM 0070, BGM 1156, BGM 0210, BGM 0195, BGM 1136, BGM 1294, BGM 0822, BGM 0279, BGM 0286, BGM 0878, BGM 0962, BGM 1306) foi de 0,6069. O ideal é que pelo menos esta medida se mantenha e não diminua e neste caso, com a retirada dos acessos duplicados, ela se

manteve. Portanto, podemos seguramente eliminar os acessos tidos como redundantes sem ocasionar perda da variabilidade genética dentro do banco de germoplasma de mandioca da Embrapa Mandioca e Fruticultura.

Embora esta metodologia talvez não seja a mais elaborada para se eliminar acessos redundantes de bancos de germoplasma, é uma metodologia mais simples e que fornece uma medida muito próxima da realidade.

A importância do uso de marcadores moleculares fica evidente, uma vez que foi detectado uma grande diferença entre os grupos separados anteriormente apenas com base nos descritores morfológicos e posteriormente com o uso de marcadores em nível de DNA e a análise conjunta. A possibilidade de unir os dados morfológicos e moleculares fornece mais confiabilidade nas análises, mostrando que todo o conjunto de informação pode ser utilizado.

CONCLUSÕES

- O uso de marcadores RAPD e SSR foi eficiente na identificação de acessos duplicados pertencentes ao banco de germoplasma BAG-Mandioca da Embrapa Mandioca e Fruticultura.

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) pelo apoio financeiro do projeto.

REFERÊNCIAS

- Chiorato, AF., Carbonell, SAM., dias, ALASD., Moura, RR., Chiavegato, MB., Colombo, CA. 2006. Identification of common bean (*Phaseolus vulgaris*) duplicates using agromorphological and molecular data. *Genetics and Molecular Biology*, 29(1): 105-111.
- Cole-Rodgers, P., Smith, DW., Bosland, PW. 1997. A novel statistical approach to analyze genetic resource evaluations using *Capsicum* as an example. *Crop Science* 37: 1000-10002.
- Creste, S., Tullman-Neto, A., Silva, SO., Figueira, A. 2003. Detection of single sequence repeat polymorphism in denaturing polyacrilamide gels by silver staining. *Plant Molecular Biology Reporter* 19: 299-306.
- Cruz, CD. 2003. Programa Genes, Versão Windows (2003.0.0). UFV, Viçosa, Brasil.
- Epperson, JE., Pachio, DH., Guevara, CL. 1996. The cost of maintaining genetic resources for cassava *Manihot esculenta* Crantz. *Proc. XIII Int. Symp. on Hort. Economics* Eds. Brumfield, RG. *Acta Horticulture*, 429, ISHS.
- Fu, YB. 2006. Redundancy and distinctness in flax germplasm as revealed by RAPD dissimilarity. *Plant Genetic Resources* 4(2): 117-124.
- Fu, YB, Guerin, S., Peterson, GW, Carlson, JE., Richards, KW. 2003. Assessment of bulking strategies for RAPD analyses of flax germplasm. *Genetic Resources and Crop Evolution* 50: 743-746.
- Gower, 1971 J.C. Gower. 1971. A general coefficient of similarity and some of its properties, *Biometrics* 27: 857-874.
- Hintum, TV. 2003. Molecular characterization of a lettuce germplasm collection. *Eucarpia Leafy Vegetables 2003* (eds. van Hintum, Th. JL., Lebeda, A., Pink, D., Schut, JW.). 99-104pp.
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull. Soc. Vaudoise Sci. Nat.* 37: 547-579.

- Liu, K., Muse, SV. 2005. PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics* 21 (9): 2128-2129.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Tamura K, Dudley J, Nei M & Kumar S (2007) *MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0*. *Molecular Biology and Evolution* 24:1596-1599.
- Virk, PS., Newbury, HJ., Jackson, MT., Ford-Lloyd, BV. 1995. The identification of duplicate accessions with rice germplasm collection using RAPD analysis. *Theoretical and Applied Genetics* 90: 1049-1055.
- Williams JGK, Kubelik AR, Livak KJ (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. ***Nucleic Acids Research* 18: 6531-6535.**