

## SELEÇÃO DE SNP (*Single-Nucleotide Polymorphism*) INFORMATIVOS PARA INTEGRAÇÃO DE ANÁLISES MOLECULARES EM MANDIOCA

Eder Jorge de Oliveira<sup>(1)</sup>, Vanderlei da Silva Santos<sup>(1)</sup>, Cláudia Fortes Ferreira<sup>(1)</sup>, Gilmara Alvarenga Fachardo Oliveira<sup>(2)</sup>, Maiana Suzarte da Silva<sup>(2)</sup>, Carlos Ivan Aguilar-Vildoso<sup>(3)</sup>

<sup>(1)</sup>Pesquisador da Embrapa Mandioca e Fruticultura, Rua da Embrapa s/n, CP007, Cruz das Almas (BA), E-mail: eder@cnpmf.embrapa.br, vssantos@cnpmf.embrapa.br, claudiaf@cnpmf.embrapa.br;

<sup>(2)</sup>Estudante de Biologia da Universidade Federal do Recôncavo da Bahia, Campus de Cruz das Almas (BA), E-mail: gfachardo@yahoo.com.br, maisuzarte@yahoo.com.br; <sup>(3)</sup>Bolsista PNPd CNPq/Embrapa Mandioca e Fruticultura, E-mail: vildoso@hotmail.com

### Introdução

O Brasil é considerado o principal centro de diversidade do gênero *Manihot* e a Amazônia, o provável centro de origem da espécie *Manihot esculenta* Crantz (Allem, 1994). Por isso, no Brasil observa-se imensa variabilidade genética, que por sua vez reflete em uma grande diversidade de usos, tanto das raízes quanto da parte aérea. A manutenção desta variabilidade em bancos de germoplasma é de importância fundamental, visando, por um lado, manter em instituições de pesquisa a variabilidade existente na natureza, de modo a evitar a erosão genética provocada pela ação humana, e por outro, fornecer a matéria prima para a obtenção de novos produtos e tecnologias.

No Brasil, existem diversos Bancos Ativos de Germoplasma (BAG) e coleções de Mandioca distribuídas em diversas instituições. Este germoplasma apresenta suficiente grau de variabilidade para atender a demanda dos melhoristas. Contudo, é preciso avançar nas atividades relacionadas à caracterização e avaliação deste importante germoplasma. Além disso, é necessário desenvolver um banco de dados, cujas informações possam ser comparadas em relação aos acessos armazenados, para evitar duplicações não intencionais, e gasto desnecessários de recursos humanos físicos e financeiros na conservação e disponibilização do germoplasma para fins de pesquisa.

Atualmente o uso de análises moleculares tem permitido avanços na caracterização de acessos de mandioca com uso de diversos tipos de marcadores a exemplo de microssatélites (*Simple Sequence Repeat* - SSR), AFLP (*Amplified Fragment Length Polymorphism*), RFLP (*Restriction Fragment Length Polymorphism*), RAPD (*Random Amplified Polymorphic DNA*) e derivados. Entretanto, uma das limitações destes tipos de marcadores é o baixo polimorfismo encontrado em indivíduos relacionados (Kawuki et al., 2009) e a grande dificuldade de integração das informações. Por outro lado, os marcadores do tipo SNP (*Single Nucleotide Polymorphisms*) representam a mais importante forma natural de variação genética (Kruglyak, 1997) e, portanto provem suficiente polimorfismo para

discriminar indivíduos altamente relacionados, além de permitirem uma automação completa da genotipagem com resultados comparáveis entre diferentes laboratórios. O objetivo deste trabalho foi analisar e selecionar marcadores SNPs com alto conteúdo informativo, de forma a contribuir para a consolidação do uso desta metodologia.

## **Material e Métodos**

### ***Material genético e extração do DNA***

Foram avaliados 1280 acessos de germoplasma do BAG-Mandioca da Embrapa Mandioca e Fruticultura. O DNA das amostras foi extraído, a partir de folhas jovens, utilizando o protocolo com CTAB (brometo de cetiltrimetilamônio) descrito por Doyle & Doyle (1990). A quantificação do DNA foi realizada após eletroforese (3 Volts/cm) de alíquotas de cada amostra, comparando-as com uma série de concentrações conhecidas de DNA (20, 40, 60, 80 e 100 ng) do fago Lambda (Invitrogen), realizada em géis de agarose a 1,0% (p/v). A concentração de DNA foi estimada por meio da comparação visual das intensidades das bandas, reveladas pela coloração com brometo de etídeo (1,0 µg/mL).

### ***Genotipagem das amostras***

A genotipagem dos SNP foi realizada utilizando a plataforma Sequenom iPLEX MassARRAY de acordo com instruções do fabricante, ou seja, com base na extensão de iniciadores alelo-específicos e em resolução por espectrometria de massa (Sequenom, San Diego, California, USA. <http://www.sequenom.com/>). Apenas amostras e SNPs que apresentaram menos de 10% de dados perdidos foram analisados.

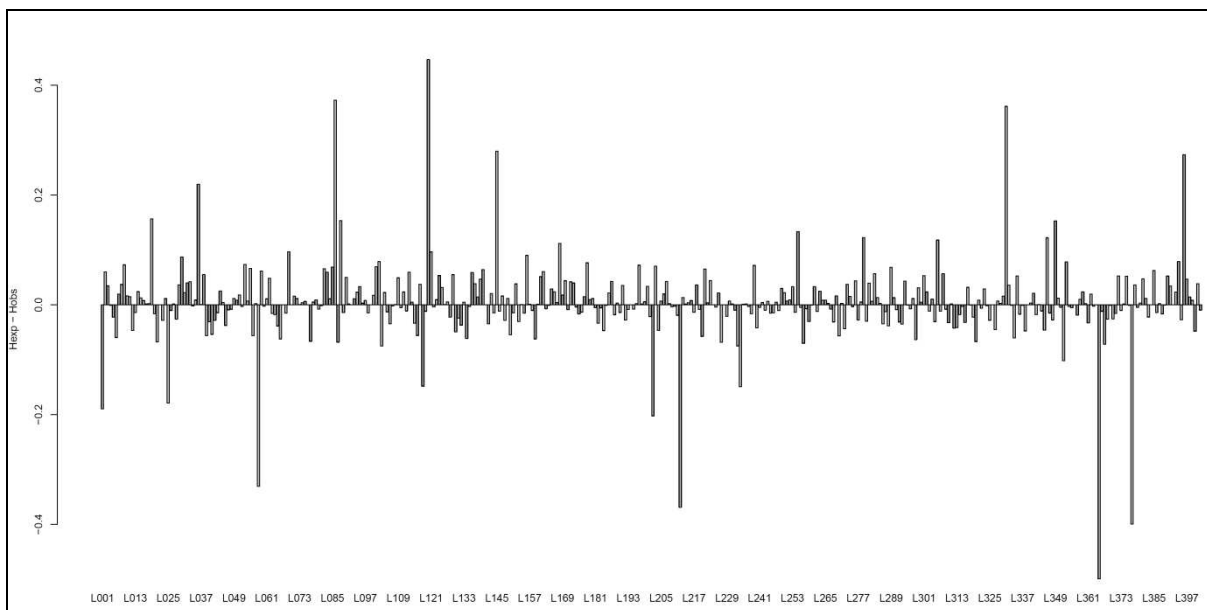
### ***Análise dos dados***

As estimativas que caracterizam o polimorfismo dos locos, como número de alelos por loco (A), heterozigiosidade observada ( $H_O$ ) e esperada ( $H_E$ ) e conteúdo informativo de polimorfismo (PIC) foram obtidas com auxílio do programa PowerMarker v.3.25 (Liu & Muse, 2005).

## **Resultados e Discussão**

Dos 402 SNPs avaliados, apenas seis foram monomórficos na análise de 1280 acessos de mandioca, o que indica alto potencial de exploração deste tipo de marcador nas análises genéticas de mandioca. Por outro lado, 226 SNPs apresentaram desvios significativos entre a proporção de indivíduos heterozigotos observados e esperados de acordo com o equilíbrio de Hardy-Weinberg ( $p < 0,05$ ). Devido ao fato de que a propagação predominante da mandioca é via clonal, e que o equilíbrio de Hardy-Weinberg somente é atingido em populações panmíticas, desvios nesta relação são esperados na cultura da mandioca.

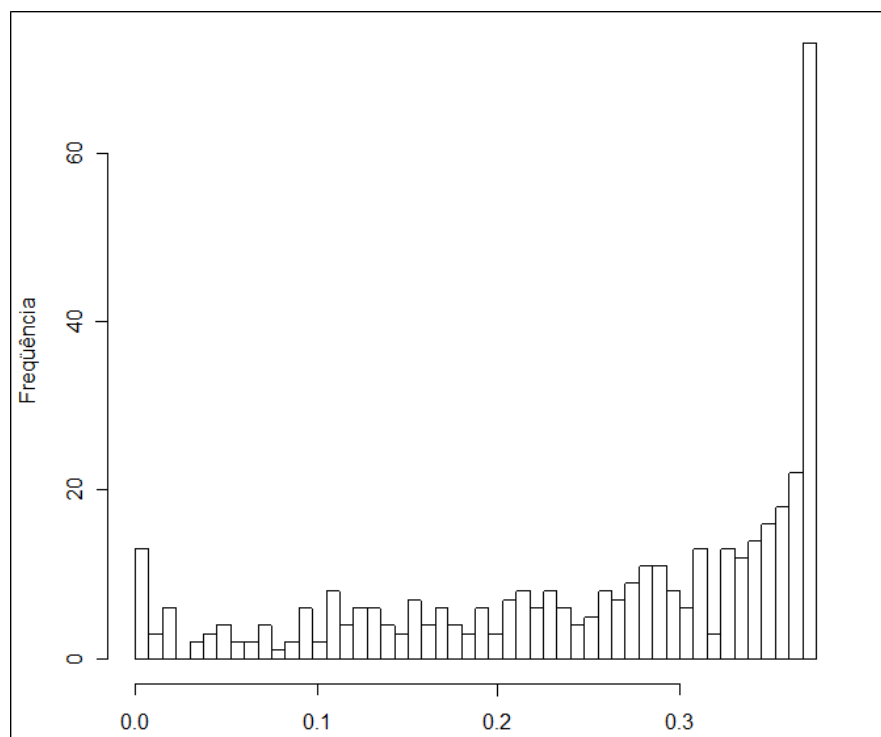
Quatro locos (SNP-23076, SNP-35735, SNP-MH076\_J24 e SNP-MH015\_B12) apresentaram  $H_O$  acima de 0,80. Entretanto, de modo geral observa-se predominância de maior  $H_O$  em relação à  $H_E$  (Figura 1). A maioria das cultivares de mandioca são oriundas de seleções em populações segregantes, ou mesmo de seleções de plantas atípicas em áreas de produção, pelos próprios produtores. Assim, em ambas as situações os mecanismos de geração (mutação) e ampliação da variabilidade (recombinação) estão presentes, de tal forma que haja alto nível de heteroziguidade observável.



**Figura 1.** Distribuição das diferenças entre as heteroziguidade esperadas e observadas na análise de 1280 acessos de mandioca com 402 marcadores do tipo SNP.

Além da heteroziguidade, o conteúdo de informação de polimorfismo (PIC) permite quantificar o polimorfismo genético dos locos em análise. Este conhecimento é importante, pois permite a otimizar o número de marcadores a serem utilizados e priorizados nos mais diversos estudos moleculares, desde análises de paternidade, estruturação de variabilidade genética, filogenia, bem como nos trabalhos de seleção assistida, mapeamento de QTLs e associativo, pois maiores valores de PIC estão associados a maior conteúdo de informação de ligação (Liu, 1997).

Os valores de PIC observados na análise dos SNPs em mandioca variaram de 0,0 a 0,37, com média de 0,26 (Figura 2). Como os SNPs são marcadores bialélicos, o valor máximo de PIC é de 0,50 (quando ambos os alelos têm a mesma frequência). Isto faz com que o PIC seja menor em comparação com outros marcadores multialélicos como os microssatélites, seja derivados de regiões genômicas aleatórias (Fregene et al., 2003) ou mesmo de regiões gênicas (Raji et al., 2009). Contudo, avanços tecnológicos na genotipagem de SNPs possibilitam alto desempenho e acurácia na geração de dados, aliado ao baixo custo para prospecção e caracterização destes marcadores, que os tornam mais competitivos para estudos de seleção assistida como ferramenta para geração de novos produtos.



**Figura 2.** Histograma da distribuição do conteúdo informativo por loco (PIC) dos marcadores SNPs, avaliados em 1280 acessos de mandioca.

Neste sentido, 196 marcadores do tipo SNPs (36328, 2496, 31596, 29001, 31240, MH006\_D12, 29279, 43443, 43129, 24971, 22343, MH131\_K24, 8206, 30907, 23496, 6076, 12955, 7391, 6780, 7557, 12301, 31229, 24399, 29465, MH151\_O08, 18552, 24056, 24341, 23199, 35630, 25143, 30029, 7259, 6497, 35831, 24808, 37493, 22991, MH144\_G13, 24468, 13969, 2554, MH025\_K13, 25091, 36667, 1979, 29172, 2100, 22334, 1167, MH077\_N08, MH059\_M24, 42475, 6464, MH138\_H14, 37791, 2257, 29599, 35279, 22768, 19186, 29686, 38236, MH028\_E01, 38161, 22620, 35514, 29934, 30590, 36107, 38, 24758, 19501, 27, 7034, 36097, 37802, 31709, 13479, 29795, MH004\_P22, 35778, 41927, 30329, 43354, 24191, 24145, 22400, 31506, 6104, 18131, 7, 9, 30749, 42312, 42961, 36722, 30719, 7092, 24939, 42768, 29065, 23460, 24672, 37661, 35449, 43844, MH170\_D24, 42694, 35873, 22647, 43055, 43569, 23966, 37850, 567, 31792, 30275, 43998, 23911, 262, 317, 37165, 29515, 22412, 12752, 35285, 23871, 30805, 30917, 31042, 30698, 35256, MH089\_P01, 23056, 24120, 2111, 36118, 36087, 23635, 31174, 23003, 7633, 7138, 35735, 37317, 29095, MH143\_F01, 30580, 23392, 28909, 618, MH067\_C15, 19312, 43376, 30411, 23076, 19231, 36634, 31001, 22633, MH098\_F13, 23341, 22972, 43649, 36278, 6889, MH148\_L01, 31802, 43227, 37155, 42823, 23725, 43005, 7823, 30333, 42209, 30634, 35878, 524, 35237, 7622, 29733, 359, 30231, 569, 29349, MH076\_J24, 30707, 30828, 6331, 42147, 25012, 37633, 29546, 35347, MH015\_B12) apresentaram PIC acima de 0,30 (até 0,37). Assim, com o avanço na descoberta de SNPs e nas tecnologias de genotipagem em larga escala prevê-se que estes marcadores desempenharão um papel cada vez mais importante nas pesquisas em genética e melhoramento da mandioca.

### Conclusões

Identificou-se 196 marcadores SNPs com alto conteúdo polimórfico para a realização de diversos estudos genéticos na cultura da mandioca, sobretudo para a estruturação da variabilidade genética, com resultados comparáveis entre as instituições detentoras de bancos de germoplasma.

### Agradecimentos

Ao CNPq e Fapesb pelo apoio financeiro.

### Referências

- ALLEM, A.C. The origin of *Manihot esculenta* Crantz (Euphorbiaceae). **Genetic Resources and Crop Evolution**, v.41, p.133–150, 1994.
- DOYLE, J.J.; DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**, v.12, p.13-15, 1990.
- FREGENE, M.A.; SUÁREZ, M.; MKUMBIRA, J.; KULEMBEKA, H.; NDEDYA, E.; KULAYA, A.; MITCHEL, S.; GULLBERG, U.; ROSLING, H.; DIXON, A.G.O.; KRESOVICH, S. Simple sequence repeat (SSR) diversity of cassava (*Manihot esculenta* Crantz) landraces: Genetic structure in a predominantly asexually propagated crop. **Theoretical and Applied Genetics**, v.107, p.1083–1093, 2003.
- KAWUKI, R.S.; FERGUSON, M.; LABUSCHAGNE, M.; HERSELMAN, L.; KIM, D-J. Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). **Molecular Breeding**, v.23, p.669–684, 2009.
- KRUGLYAK, L. The use of a genetic map of biallelic markers in linkage studies. **Nature Genetics**, v.17, p.21–24, 1997.
- LIU K, MUSE SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. **Bioinformatics Applications Note**, v.21, p.2128–2129.
- LIU, B.H. Statistical genomics: linkage, mapping, and QTL analysis. Boca Raton: CRC Press, 1997. 611 p.
- RAJI, A.A.; ANDERSON, J.V.; KOLADE, O.A.; UGWU, C.D.; DIXON, A.G.; INGELBRECHT, I.L. Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, Polymorphisms, and cross-taxa utility. **BMC Plant Biology**, 2009. doi:10.1186/1471-2229-9-118.