

**UNIVERSIDADE FEDERAL DO AMAZONAS**  
**PROGRAMA MULTI-INSTITUCIONAL DE PÓS-GRADUAÇÃO EM**  
**BIOTECNOLOGIA**

**ANA MARA OLIVEIRA DA SILVA**

**DIVERSIDADE E ESTRUTURA GENÉTICA DA COLEÇÃO**  
**REGIONAL DE GERMOPLASMA DE MANDIOCA DA EMBRAPA**  
**AMAZÔNIA OCIDENTAL**

**MANAUS**

**2014**

**DIVERSIDADE E ESTRUTURA GENÉTICA DA COLEÇÃO  
REGIONAL DE GERMOPLASMA DE MANDIOCA DA EMBRAPA  
AMAZÔNIA OCIDENTAL**

**ANA MARA OLIVEIRA DA SILVA**

Bacharel em Biotecnologia

Orientador: Dr. **CHARLES ROLAND CLEMENT**

Co-orientadora: Dra. **NELCIMAR REIS SOUSA**

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia, Universidade Federal do Amazonas, como parte dos requisitos para obtenção do título de Mestre em Biotecnologia.

**MANAUS**

**2014**

Ficha Catalográfica  
(Catalogação realizada pela Biblioteca Central da UFAM)

Silva, Ana Mara Oliveira da

S586d Diversidade e estrutura genética da coleção regional de germoplasma de mandioca da EMBRAPA Amazônia Ocidental / Ana Mara Oliveira da Silva, 2014.

47f. il. color.

Dissertação (mestrado em Biotecnologia) – Universidade Federal do Amazonas.

Orientador: Prof. D<sup>r</sup>. Charles Roland Clement

Co-orientador: Dra. Nelcimar Reis Sousa

1. Mandioca 2. Germoplasma vegetal 3. Diversidade genética I. Clement, Charles Roland (Orient.) II. Universidade Federal do Amazonas III. Título

CDU(2007) 633.493(811)(043.3)

**ANA MARA OLIVEIRA DA SILVA**

**DIVERSIDADE E ESTRUTURA GENÉTICA DA COLEÇÃO  
REGIONAL DE GERMOPLASMA DE MANDIOCA DA  
EMBRAPA AMAZÔNIA OCIDENTAL**

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia, Universidade Federal do Amazonas, como parte dos requisitos para obtenção do título de Mestre em Biotecnologia.

Aprovado em 09 de julho de 2013.

**BANCA EXAMINADORA**

Prof<sup>a</sup> Dra. Doriane Picanço Rodrigues, Presidente  
Universidade Federal do Amazonas

Dra. Elisa Ferreira Moura, Membro  
Embrapa Amazônia Oriental

Dra. Nelcimar Reis Sousa, Membro  
Embrapa Amazônia Ocidental

## AGRADECIMENTOS

À Deus acima de tudo por ser o autor da vida e iluminar e abençoar todos os meus passos.

À todos os meus familiares, em especial minha mãe e meu pai, meus irmãos e sobrinhos, à minha sobrinha Yamina por quem tenho muito apreço.

Ao meu orientador Prof. Dr. Charles Clement pela oportunidade, ensinamentos, orientação, objetividade, dedicação e acima de tudo pela paciência com seus orientados.

À Universidade Federal do Amazonas por meio do Programa de Pós-graduação em Biotecnologia, pela oportunidade para que eu realizasse o mestrado e à todo corpo docente do PPGBIOTEC-UFAM, pela competência e qualidade de ensino.

Ao coordenador do PPGBIOTEC-UFAM Prof. Dr. Edmar Andrade, pelos ensinamentos e amizade ao longo desses dois anos de caminhada, obviamente pelos ensinamentos e disponibilidade em atender as necessidades dos discentes do PPGBIOTEC.

Ao CNPq pela concessão da bolsa de estudos e pelo financiamento de minha pesquisa, importantíssimo para a realização deste trabalho.

À Embrapa Amazônia Ocidental pelo suporte à pesquisa.

Em especial, aos amigos e funcionários do Laboratório de Biologia Molecular, a todos os bolsistas que já fizeram pesquisa, àqueles que estão começando ou concluindo que ajudaram direta e indiretamente, em especial ao Jeferson pela paciência e disponibilidade de ajudar principalmente quando mais precisamos.

Ao Dr. Gilvan Ferreira pela ajuda, ensinamentos, pelas tantas vezes que se disponibilizou a sacrificar sábados, domingos e feriados, até altas horas no laboratório, tudo em prol da ciência.

À Dra. Nelcimar minha co-orientadora, que me acompanha desde a graduação, pela oportunidade, disposição em ajudar, pela amizade, conhecimento que sempre se dedicou a transmitir, pelo exemplo como cientista e pela paciência na etapa final do trabalho quando já não conseguimos nem dizer “bom dia” para o orientador.

A todos os meus amigos, Jefferson Ferreira que sempre esteve ao meu lado desde a graduação, um irmão pra mim, ao nosso grupo de trabalho, Saulo, Adriel e José.

À Ana Cláudia Gama pela eterna amizade, à Karina Bichara pela amizade e à Thyanne Bastos pela ajuda, amizade e companheirismo em todos os momentos.

A todos que direta e indiretamente participaram deste trabalho e de minha formação.

## RESUMO

A mandioca (*Manihot esculenta* Crantz) desempenha importante papel social como fonte de alimentação nas regiões tropicais mundiais. O germoplasma conservado constitui a base para aproveitamento tecnológico da espécie no desenvolvimento de novas cultivares. Para que seja eficientemente utilizado necessita de diferentes métodos de caracterização genética. Os marcadores com base em elementos transponíveis são indicados para análises genéticas devido as suas qualidades de reprodutibilidade e polimorfismo abundante. Foram utilizados 430 acessos de mandioca que compõem o germoplasma da Embrapa, coletados na bacia amazônica. As sequências de retrotransposons foram localizadas no banco de dados do Phytozome, os primers IRAP desenhados com o programa Primer3, a estruturação dentro do germoplasma foi detectada através do software STRUCTURE v 2.2 e a diversidade genética foi avaliada utilizando o software Popgen v1.32. Para os dados de IRAP o software STRUCTURE sugeriu a existência de dois agrupamentos genéticos (K=2) um com 93 e outro com 127 variedades do total de 430 plantas, considerando apenas a fidelidade acima de 80%. A AMOVA para K = 2 revelou mais variação dentro do grupo (89%) que entre os grupos (11%). Os seis pares de primers IRAP foram informativos para avaliar a diversidade genética, com médias de 96% de polimorfismo, 0.4 de heterozigosidade e com índice de Shannon de 0.57, porém, não detectaram estruturação dentro do germoplasma de mandioca comparada com a estruturação assistida por outros marcadores.

**Palavras-chave:** IRAP; Polimorfismo; Retrotransposon; Diversidade Genética

## ABSTRACT

Manioc (*Manihot esculenta* Crantz) performs an important social role as a food source in the tropical world. The germplasm maintained manioc forms the basis for technological exploitation of the species in the development of new cultivars. To be efficiently used require different methods of genetic characterization. Markers based on transposable elements are suitable for genetic analyzes because of their qualities of reproducibility and abundant polymorphism. We used 430 manioc accessions that compose the Embrapa germplasm collected in the Amazon basin. The retrotransposons sequences were located in the Phytozome database, IRAP primers designed with Primer3 program, structuring within the germplasm has been detected by STRUCTURE software v. 2.2 and the genetic diversity was assessed using the Popgen software v1.32. IRAP data to the software STRUCTURE suggested the existence of two gene clusters ( $k = 2$ ) with 93 and other with 127 varieties of total 430 plants, considering only the fidelity above 80%. The AMOVA for  $K = 2$  showed greater variation within the group (89%) than among groups (11%). The six IRAP primer pairs were informative for assessing genetic diversity, with averages of 96% polymorphism, 0.4 heterozygosity and Shannon index 0.57, however, did not detect structuring within cassava germplasm compared to other assisted structuring markers.

**Key words:** IRAP; polymorphism; retrotransposon; Genetic Diversity

## LISTA DE FIGURAS

<b>Figura 1</b> – Banco de germoplasma de cultivares de mandioca da Embrapa Amazônia Ocidental .....	14
<b>Figura 2</b> – Elementos Transponíveis de classe I, sem LTR .....	19
<b>Figura 3</b> – Retrotransposons do Tipo LTR, classe II .....	20
<b>Figura 4</b> – Organização geral das LTRs de um retrotransposon da superfamília copia .....	21
<b>Figura 5</b> – Banco de germoplasma de mandioca da Embrapa Amazônia Ocidental .....	24
<b>Figura 6</b> – Perfil de amplificação e polimorfismo observado com primer IRAP ME_8 em variedades do BAG de germoplasma de mandioca .....	28
<b>Figura 7</b> – Valor de $\Delta K$ de possíveis agrupamentos de 430 variedades de mandioca brasileira obtidos de dez simulações de análise da estrutura .....	28
<b>Figura 8</b> – Agrupamento dos 220 acessos de mandioca de acordo com um algoritmo baseado no modelo Bayesiano implementado no software Structure .....	29
<b>Figura 9</b> – Agrupamento dos acessos de mandioca de acordo com um algoritmo baseado em modelo Bayesian implementado no programa Structure para ( $K = 3$ ) com fidelidade acima de 80% .....	30

## LISTA DE TABELAS

<b>Tabela 1</b> – Características de primers IRAP para <i>Manihot esculenta</i> .....	27
<b>Tabela 2</b> – Média da análise de variância molecular (AMOVA) para agrupamentos por marcadores IRAP para grupos de variedades de mandioca da Bacia Amazônica estimadas no software STRUCTURE .....	30
<b>Tabela 3</b> – Níveis de polimorfismo e diversidade genética detectados com marcadores IRAP entre mandiocas usadas nesse estudo.....	31

## SUMÁRIO

<b>JUSTIFICATIVA</b> .....	<b>10</b>
<b>INTRODUÇÃO</b> .....	<b>11</b>
1.1 Mandioca .....	11
1.1.1 Domesticação e evolução .....	11
1.1.2 Variabilidade genética e Banco de germoplasma.....	12
1.2 Marcadores moleculares .....	14
1.2.1 Marcadores Moleculares Baseados em Sequências Repetitivas .....	15
1.2.1.1 Sequências Repetitivas em tandem.....	15
1.3 Elementos transponíveis .....	16
1.3.1 Elementos de Classe II .....	18
1.3.2 Elementos de Classe I.....	18
1.3.2.1 Retrotransposons sem LTR .....	19
1.3.2.2 Retrotransposons com LTR.....	19
1.4 Marcador baseado em Retrotransposon – IRAP (Inter-Retrotransposon Amplified Polymorphism) .....	21
2. OBJETIVOS .....	23
2.1 OBJETIVO GERAL.....	23
2.2 OBJETIVOS ESPECÍFICOS .....	23
3. MATERIAIS E MÉTODOS.....	24
3.1 Material Vegetal .....	24
3.2 Desenvolvimento de Marcadores para IRAP e condições de PCR .....	24
3.3 Análise dos Dados .....	25
3.3.1 Estrutura Genética das Populações.....	25
3.3.2 Diversidade Genética com marcadores IRAP .....	26
4. RESULTADOS .....	27
Desenho de marcadores para IRAP .....	27
Estrutura Genética em mandioca com marcadores IRAP .....	28
Diversidade Genética dentro do germoplasma de mandioca.....	31
5. DISCUSSÃO .....	33
<b>CONCLUSÕES GERAIS</b> .....	<b>34</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>35</b>
<b>ANEXO</b> .....	<b>42</b>

## JUSTIFICATIVA

A mandioca (*Manihot esculenta* Crantz) é um dos principais cultivos em países tropicais e subtropicais. Na Amazônia, as populações tradicionais a consideram uma cultura de grande importância, por isso a seleção e melhoramento ainda são práticas rotineiras entre os agricultores. A interação entre melhoristas tradicionais e científicos é um fator fundamental para o sucesso de qualquer programa de melhoramento genético. O intercâmbio entre agricultores e empresas de pesquisa, como a Embrapa, por exemplo, possibilitaria o enriquecimento das coleções de germoplasma para serem usados em programa de melhoramento e aproveitamento tecnológico da espécie, beneficiando tanto os agricultores quanto a Embrapa.

A Embrapa Amazônia Ocidental mantém uma coleção de germoplasma de mandioca desde 1994 com cerca de 500 acessos coletados em diferentes localidades da bacia amazônica. Para a utilização do germoplasma no melhoramento torna-se necessário priorizar a caracterização genética e estabelecer coleções representativas da diversidade. A caracterização genética permite a identificação de acessos duplicados, reduzindo o desperdício de tempo e recursos financeiros, e as redundâncias geradas por sinonímias entre as cultivares. O uso de marcadores moleculares se apresenta como uma das principais alternativas para ampliar o conhecimento do germoplasma conservado.

Os bancos de dados de sequências de DNA têm sido explorados para identificação de marcadores de terceira geração com base em retroelementos IRAP (*Inter-Retrotransposon Amplified Polymorphism*), pois esses marcadores são considerados atrativos pela qualidade de sua reprodutibilidade e polimorfismo abundante e estão se tornando uma ferramenta para aprender sobre a evolução do genoma e função dos genes nas espécies.

O melhoramento genético de qualquer espécie depende da diversidade genética disponível, de informações sobre essa diversidade e, sobretudo, da relação entre o DNA e os fenótipos importantes para os melhoristas, produtores e consumidores. As informações geradas nesta pesquisa trarão benefícios à comunidade científica, ampliando o conhecimento e permitindo a identificação de novos materiais genéticos, à Embrapa, organizando a coleção de germoplasma e possibilitando o melhoramento da espécie, e por fim, mas não menos importante, aos agricultores com desenvolvimento de novas cultivares mais produtivas.

## INTRODUÇÃO

### **1.1 Mandioca**

A mandioca é uma dicotiledônea perene, arbustiva, pertencente à família das Euforbiáceas, subfamília Crotonoideae, cujo gênero *Manihot* possui cerca de 100 espécies (Alves, 2002). No gênero *Manihot* duas espécies apresentam importância econômica: a *Manihot esculenta* Crantz (mandioca) com raízes tuberosas para produção de farinha, amido, consumo *in natura*, uso industrial, e *Manihot glaziovii* Muell. Arg. (maniçoba) para produção de látex. A única espécie cultivada comercialmente é a *Manihot esculenta* Crantz (Faraldo, 2000).

A mandioca é frequentemente cultivada em roças de agricultura autóctone ou tradicional, por apresentar bom desenvolvimento em solos pobres, resistência a pragas e doenças e adaptação às mais diversas condições edafoclimáticas (Hershey, 1992). Sua importância social e econômica consiste basicamente em ser a principal fonte de carboidratos e de subsistência para cerca de um bilhão de pessoas no mundo, é usada para alimentação animal e ainda apresenta aplicação industrial como biocombustível (Balat e Balat, 2009).

O continente africano é o maior produtor mundial de mandioca com uma média de 52% nos últimos seis anos, seguido pela Ásia (32,6%), Américas (14,5%) e Oceania (0,8%). O Brasil é um dos maiores produtores de mandioca do mundo, sendo que, noventa e cinco por cento das propriedades produtoras possuem bases familiares.

A região Norte do Brasil se destaca como a segunda principal produtora de mandioca alternando como maior produtor do país com a região Nordeste. Conforme os dados do Instituto Brasileiro de Geografia e Estatística (IBGE), a região Norte foi produtora de 29,29% da produção total do país em 2012, a região com a maior produção foi a nordeste com uma representatividade de 32,6% e um total de 8,2 milhões de toneladas. Diante da grande produção de mandioca no norte do Brasil, passa a ser a tuberosa um fator importante para a região, pois gera renda e segurança alimentar.

#### **1.1.1 Domesticação e evolução**

A domesticação pode ser considerada como uma classe de evolução em que os humanos adicionam suas ações à ação da seleção natural, às vezes em consonância com esta, às vezes em dissonância (Clement, 2001). A domesticação tem sido chamada de evolução dirigida e evolução prática (Harlan, 1992), pois a evolução natural melhora a adaptação de uma população a seu meio, enquanto que a domesticação adapta uma população às necessidades humanas.

Para Clement (1999) a domesticação de plantas é definida como um processo co-evolucionário em que a seleção humana, inconsciente e consciente, nos fenótipos de populações de plantas promovidas, manejadas ou cultivadas resulta em mudanças nos genótipos das populações que as tornam mais úteis aos humanos e melhor adaptadas às intervenções humanas no ambiente.

Há dois centros de diversidade de espécies do gênero *Manihot*. A maioria das espécies ocorre no norte da América do Sul (~ 80 espécies), e um centro secundário de diversidade ocorre no México e América Central (17 espécies) (Chavarriaga-Aguirre e Halsey, 2005). O Brasil é considerado o provável centro de origem e diversificação da espécie cultivada (Allem, 2002). Da diversidade encontrada no Brasil, cerca de 4.132 acessos de mandioca foram coletados e conservados em coleções ex situ (Tomich et al., 2008).

A domesticação da mandioca ainda não está bem esclarecida, porém estudos moleculares estão contribuindo para mudar esse cenário, em geral mostrando que a mandioca foi domesticada apenas uma vez, na América do Sul, sem contribuição do *pool* de espécies da Mesoamérica (Roa et al., 1997, 2000; Olsen e Schaal, 1999). Olsen e Schaal (1999) indicaram, a partir de análises moleculares, que a mandioca foi domesticada especificamente de populações de *M. esculenta* subsp. *flabellifolia* ocorrendo ao longo do sul da bacia Amazônica, nos estados brasileiros do Acre, Rondônia e Mato Grosso e provavelmente se estendia a Bolívia.

O estudo desta dinâmica evolutiva sob domesticação é essencial para se entender a cultura da mandioca e planejar sua conservação, pois diversos fatores de ordem cultural, social e econômica podem resultar em erosão genética severa da base genética local (Peroni e Hanazaki, 2002; Emperaire e Peroni, 2007).

### **1.1.2 Variabilidade genética e Banco de germoplasma**

A cultura da mandioca apresenta uma grande variabilidade genética com um elevado número de variedades disponíveis. Apesar de, ao longo do tempo, ter sido propagada vegetativamente por interferência humana, manteve a reprodução sexuada ativa, promovendo a amplificação da variabilidade genética (Silva et al., 2001). Dentro da espécie *Manihot esculenta* já foi identificada diversidade genética para quase todos os caracteres estudados, incluindo aqueles de natureza morfológica, agrônômica, de resistência às principais pragas e doenças e de qualidade.

Nos diversos ambientes onde a mandioca se diversificou, a seleção resultou numa ampla diversidade de clones, com adaptação específica a determinados ecossistemas (Hershey, 1988).

Em função da grande importância da cultura da mandioca no Brasil foram criados e vêm sendo mantidos bancos de germoplasma. Estes têm como finalidade principal reunir em um local, parte da variabilidade genética (germoplasma) visando evitar a perda de genes ou erosão genética, para dessa forma assegurar ampla base genética aos programas de melhoramento (Fukuda et al., 1996). Os bancos de germoplasma, normalmente, são constituídos por variedades tradicionais (etnovarietades), variedades modernas (melhoradas) e variedades silvestres do mesmo gênero da cultura (Vieira et al., 2008).

A caracterização e avaliação do germoplasma de mandioca tornam-se fundamentais para a utilização mais eficiente da coleção, pois, possibilitam a identificação de genótipos com características superiores e herdáveis.

A caracterização do germoplasma é realizada por meio de dados de passaporte, caracteres fenotípicos (agronômicos e descritores morfológicos) e marcadores moleculares. A utilização de marcadores moleculares na caracterização de germoplasma de mandioca vem crescendo nos últimos anos, principalmente em função de estes não sofrerem influência do ambiente e possibilitarem a geração de uma grande quantidade de informação referente ao genoma da espécie (Mühlen et al., 2000). No entanto, somente são úteis ao melhorista se forem acompanhados de caracterização morfo-agronômico, e serão mais úteis ainda, se existir uma relação entre os dois tipos de caracteres (molecular e morfo-agronômico), que é de fundamental importância para a determinação da variabilidade genética nos bancos de germoplasma e no planejamento de programas de melhoramento genético.

No Brasil criou-se uma rede nacional composta por seis bancos regionais de germoplasma de mandioca todos os quais vêm priorizando a caracterização. O Banco da região dos Tabuleiros Costeiros, localizado na Embrapa Mandioca e Fruticultura Tropical, que coordena os outros Bancos Ativos de Germoplasma de Mandioca, conta com 1.720

acessos. Constituem a sub-rede de Bancos Ativos de mandioca no País, os seguintes: o BAG da Embrapa Cerrados, com 445 acessos; o BAG da Embrapa Amazônia Ocidental, com 430 acessos; o BAG da Embrapa Amazônia Oriental, com 363 acessos; o BAG da Região do Semi-Árido conta com 415 acessos. O BAG da região do subtropical, mantido pela Empresa Catarinense de Pesquisa (Epagri), conta com 1.062 acessos (Nass, 2006).

A coleção de Germoplasma da Embrapa Amazônia Ocidental existe desde 1994, os genótipos de mandioca que o constituem foram coletados em diversas localidades da bacia amazônica e alguns poucos de outras regiões, porém faz-se necessário priorizar a caracterização genética para utilizá-la posteriormente em programa de melhoramento.



**Figura 1:** Banco de germoplasma de cultivares de mandioca da Embrapa Amazônia Ocidental.

**Foto:** Nelcimar Reis, 2010.

## ***1.2 Marcadores moleculares***

Características de DNA que diferenciam dois ou mais indivíduos e são herdadas geneticamente são conhecidas como marcadores moleculares (Milach, 1998) e podem ser definidos como locos genômicos, detectados através de sonda ou iniciadores específicos que, em virtude de sua presença, distingue de forma inequívoca a característica cromossômica que ele representa bem como as regiões flaqueadoras nas extremidades 3' e 5' (Barcaccia et al., 2000).

Um marcador genético ideal deve apresentar os seguintes atributos: a) alto nível de polimorfismo, b) estabilidade em diferentes ambientes, c) detectar grande número de locos não ligados e d) ser de herança simples (Milach, 1998; Matioli, 2001).

Um marcador molecular pode ser considerado codominante (RFLP, SSR, Isoenzimas) quando permite diferenciar nos cromossomos homólogos fragmentos de um mesmo tamanho (genótipo homozigótico) ou não (genótipo heterozigótico), ou dominante (RAPD, AFLP, VNTR, IRAP), quando não permitem a distinção entre genótipos homozigóticos e heterozigóticos de maneira que o resultado da amplificação de um determinado fragmento, no gel, identifica-se pela presença ou ausência de uma banda, que neste caso seria a presença ou ausência de alelos de um mesmo loco (Lopes et al., 2002). Embora a classe de marcadores codominantes seja em geral mais informativa, os marcadores dominantes são mais fáceis de desenvolver e utilizar.

A partir do advento das técnicas modernas de biologia molecular, surgiram diversos métodos de detecção de polimorfismo genético. Inicialmente, a utilização de enzimas de restrição permitiu a análise de polimorfismo por comprimento de fragmentos de restrição de DNA, o RFLP (*Restriction Fragment Length Polymorphism*; Botstein et al., 1980). O desenvolvimento do processo de amplificação em cadeia de polimerase (PCR) (Mullis e Faloona, 1987; Saiki et al., 1988) levou à descrição de outras classes de marcadores moleculares, tais como, RAPD (*Random Amplified Polymorphic DNA*; Williams et al., 1990) e AFLP (*Amplified Fragment Length Polymorphism*; Vos et al., 1995).

As técnicas de PCR, clonagem e seqüenciamento de DNA têm possibilitado um rápido acúmulo de informações sobre a estrutura de genomas e tem contribuído para a caracterização de cultivares (Ferreira e Grattapaglia, 1998). A disseminação do seu uso contribuiu para a descoberta e estudo de diversas classes de seqüências de DNA repetitivos.

### ***1.2.1 Marcadores Moleculares Baseados em Seqüências Repetitivas***

#### ***1.2.1.1 Seqüências Repetitivas em tandem***

Sabe-se que uma grande proporção do genoma de eucariotos é composta de seqüências repetidas de DNA. Esta proporção, no entanto, é variável. Várias classes de seqüências repetidas de DNA têm sido descritas e caracterizadas em várias espécies animais e vegetais que diferem quanto ao número e composição de nucleotídeos. Por isso, do ponto de vista de marcadores moleculares, estas seqüências repetitivas tornam-se importantes

ferramentas para análise genética, porque oferecem a possibilidade de amostrar simultaneamente um grande número de locos genéticos polimórficos dispersos no genoma.

Na classe de marcadores codominantes envolvendo sequências repetitivas *em tandem* enquadram-se os SSRs (*Simple Sequence Repeats*; Litt e Luty, 1989), ou microssatélites, que consistem em pequenas seqüências com 1 a 6 nucleotídeos de comprimento que, no total, não ultrapassam geralmente o tamanho de 200. Existem em abundância em plantas e são hoje os mais populares e mais utilizados dos marcadores moleculares disponíveis (Oliveira et al., 2006).

Outra classe envolvendo sequências repetitivas *em tandem* são o VNTR (*Variable Number of Tandem Repeats*; Jeffreys et al., 1985), um loco hipervariável, também denominado minissatélite, constitui-se de seqüências idênticas repetidas lado a lado e possuem de 15 a 100 pares de bases e, em cada loco hipervariável, são repetidas até 50 vezes (Ferreira e Grattapaglia, 1998). Atualmente, esse polimorfismo pode ser detectado por PCR, utilizando *primers* que reconhecem sequências externas que rodeiam as sequências repetidas (Heat et al., 1993). Os minissatélites têm sido utilizados no melhoramento de plantas, identificação de variedades, cultivares e clones, para estudar a diversidade genética e fazer a identificação dos indivíduos (*fingerprinting*) em várias espécies (Stockton et al., 1990; Ramakrishna et al., 1995 ).

Alguns tipos de marcadores moleculares, reconhecidos como marcadores de terceira geração por alguns autores (Singh, 2011), têm sido desenvolvidos com o avanço da era genômica, pois o sequenciamento do DNA tem facilitado a identificação dos mesmos. Dentre estes, os que estão sendo bastante utilizados em plantas são os marcadores baseados em elementos transponíveis, particularmente em retrotransposon LTR, pois estes são altamente dinâmicos (Wicker et al., 2003).

### ***1.3 Elementos transponíveis***

Os elementos de transposição (TEs) ou transposons foram descobertos em milho durante a década de 1940 por Barbara McClintock, sendo estes considerados elementos genéticos móveis. A mobilidade dos transposons ocorre pela recombinação entre suas sequências de DNA posicionadas nas extremidades e uma seqüência no DNA da célula hospedeira. O deslocamento pode ocorrer com ou sem a duplicação do elemento. Em alguns casos a recombinação envolve um intermediário temporário de RNA (McClintock, 1950).

Normalmente, quando os elementos de transposição se deslocam, eles apresentam pouca seletividade de seqüência na escolha dos sítios de inserção. Embora muito raro, os transposons podem se inserir dentro de genes, prejudicando completamente a função do gene e dentro de seqüência gênicas reguladoras, onde sua presença pode provocar alterações na expressão desse gene (McClintock, 1956). Essas perdas de função e alterações na expressão gênica levaram à descoberta dos elementos de transposição. Não é de surpreender, portanto, que os elementos de transposição estejam entre as causas mais comuns de novas mutações, em diversos organismos. A capacidade de se inserir no DNA também levou à sua modificação e utilização como agentes mutagênicos e vetores de DNA na biologia experimental (Watson et al., 2006).

Os elementos de transposição estão presentes nos genomas de todas as formas de vida. As análises comparativas de seqüências genômicas revelam duas observações fascinantes. A primeira é que as seqüências relacionadas aos elementos transponíveis podem compreender enormes frações do genoma de um organismo. Por exemplo, mais de 50% dos genomas de seres humanos e do milho são compostos por seqüências de DNA relacionadas aos elementos transponíveis (Feschotte, 2004). Em um contraste extraordinário, apenas uma pequena percentagem de seqüências (< 2% em humanos) realmente codifica as proteínas celulares. Segundo, o conteúdo de elementos transponíveis em diferentes genomas é altamente variável (Watson et al., 2006).

Devido a sua habilidade de se inserir em novos genomas sem encontrar restrições, os elementos de transposição estão sujeitos a altas taxas de mutação e a uma alta pressão seletiva levando a um aumento na freqüência de elementos incompletos e não autônomos (Capy et al., 1994).

Todo o aparato utilizado para a sua transposição é sintetizado na maioria das vezes pelo próprio elemento (elementos autônomos), mas em alguns casos ele pode utilizar as enzimas sintetizadas por outros elementos para realizar a sua movimentação (elementos não-autônomos).

Duas classes de elementos transponíveis existem e se diferem quanto ao mecanismo de transposição molecular: Classe I composta pelos Retrotransposons, que se movem através do mecanismo de “copiar e colar”, utilizando um intermediário de RNA, e gerando uma nova cópia de elemento transponível que se insere em outro local no genoma. Classe II composta pelos Transposons se movem através de um mecanismo de “cortar e colar”, utilizando um intermediário de DNA, se mudam para um novo local genômico sem aumento obrigatório no número de cópias (Wicker et al., 2007).

### ***1.3.1 Elementos de Classe II***

Eles são divididos em subclasses ou famílias de acordo com a similaridade entre as suas seqüências e se encontram dispersos pelo genoma de procariotos e eucariotos.

O mecanismo de transposição utilizado pelos elementos dessa classe é o não-replicativo, onde não ocorre duplicação do transposon. O elemento transponível se desloca diretamente de um local para outro mantendo suas características, mas este tipo de movimento não favorece o aumento no número de cópias do transposon, já que ele é perdido do local original (Lewin, 2004).

Os transposons possuem seqüências que atuam como sítios de recombinação e genes que codificam proteínas que participam na recombinação. Os elementos que possuem genes para codificação da transposase (enzima responsável pela transposição) são chamados transposons autônomos. Entretanto, os genomas também apresentam muitos segmentos de DNA móveis ainda mais simples, conhecidos como transposons não-autônomos. Esses elementos possuem apenas as repetições terminais invertidas, necessárias à transposição. Em células que tenha também um transposon autônomo que codifique uma transposase, esta irá reconhecer também as repetições terminais invertidas do elemento não-autônomo, permitindo sua transposição. Entretanto, na ausência desse transposon “auxiliar” (para fornecer a transposase) os elementos não autônomos permanecem estáticos, incapazes de deslocamento (Watson et al., 2006).

### ***1.3.2 Elementos de Classe I***

Os elementos desta classe são conhecidos por Retrotransposons. Eles se movimentam através de um intermediário de RNA que é codificado para DNA por uma transcriptase reversa, produzida por eles mesmos, antes da sua nova inserção. Por sua vez, esta classe é subdividida em duas subclasses. Os retrotransposons com LTRs (*Long Terminal Repeats*) diretas ou invertidas, e os sem LTRs (Capy et al., 1994).

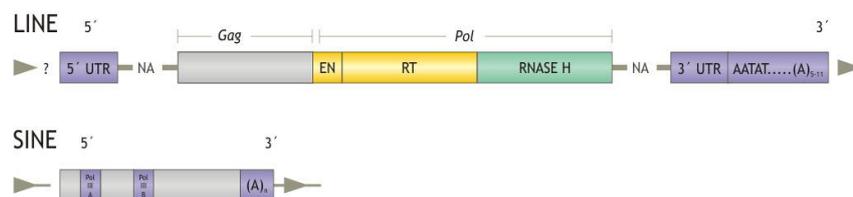
### 1.3.2.1 Retrotransposons sem LTR

Esta subclasse é dividida em duas superfamílias compostas pelos elementos *Long Interspersed Nuclear Elements* (LINEs) e *Short Interspersed Nuclear Elements* (SINEs), a diferença entre essas duas superfamílias é a presença, nos LINEs, ou ausência, nos SINEs da enzima transcriptase reversa. Tanto os elementos LINEs quanto os SINEs tem sido encontrados em plantas, fungos e animais. Eles são encontrados em grande número de cópias nos genomas eucariotos (Kumar e Bennetzen, 1999; Hansen e Heslop-Harrison, 2004).

A superfamília LINEs compreende elementos com tamanhos que variam entre 5 a 8Kb, e possuem uma seqüência rica em adeninas no término 3', muito semelhante as caudas *poli A* presente em RNAs mensageiros. Esses elementos possuem *Open-Reading Frames* (ORFs) que codificam algumas enzimas necessárias para a transposição, como a transcriptase reversa e a RNaseH (Figura 2). Vários elementos LINEs apresentam especificidade por certos sítios de inserção, como os elementos *R1 e R2* que tem preferência por uma região da seqüência do gene da subunidade maior do RNA ribossômico.

Os elementos da superfamília dos SINEs possuem tamanhos menores que 500pb. Todos os elementos desta superfamília, com exceção dos elementos *Alu*, por não apresentarem seqüências correspondentes a transcriptase reversa, são geralmente dependentes dos LINEs para realizar sua transposição. Possuem seqüências homólogas e não homólogas à RNA transportador e uma região variável e rica em adeninas e timinas (Capy et al., 1998).

#### Retrotransposons Não - LTR



**Figura 2:** Elementos Transponíveis de classe I, sem LTR.

Fonte: White et al. (1994)

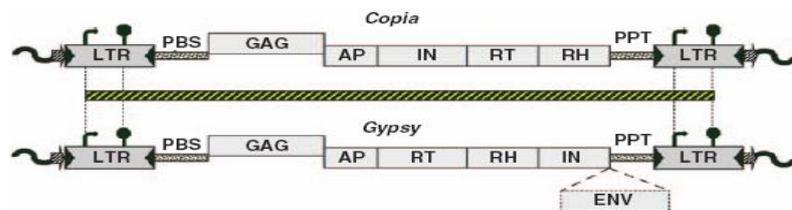
### 1.3.2.2 Retrotransposons com LTR

São elementos estruturalmente similares aos retrovírus. Possuem longas repetições nucleotídicas nas extremidades 5' e 3'. De uma maneira geral, estas repetições terminais

flanqueiam uma região central que contém três ORFs. A sequência das ORFs pode variar entre os elementos deste grupo.

A primeira ORF refere-se ao gene *gag* (*Group-specific Anti Gene*) que produz uma poliproteína que é processada em três proteínas maduras: a matriz, o capsídeo e o nucleocapsídeo. Essas três proteínas possuem similaridade ao componente do capsídeo dos retrovírus. A outra ORF constitui-se do gene *pol* que codifica as enzimas necessárias à transposição do elemento que são as enzimas protease (Pr), transcriptase reversa (TR), RNaseH e integrase (INT). A última ORF está presente em algumas famílias desta classe, podendo ou não produzir uma proteína funcional: ela corresponde ao gene *env*, que codifica a proteína do envelope viral nos retrovírus.

De acordo com a sequência codificante dentro da segunda ORF, esta subclasse de retroelementos pode ser dividida em duas superfamílias: *Ty1-copia* com a sequência 5' Pr-Int-TR-RNaseH 3', e *Ty3-gypsy* com a sequência 5' Pr-TR-RNaseH-Int 3' (Figura 3). A superfamília *Ty3-gypsy* pode ser ainda dividida em duas famílias de acordo com a presença (família *gypsy*) ou ausência (família *Ty3*) do gene *env* (Marco e Marín, 2005).

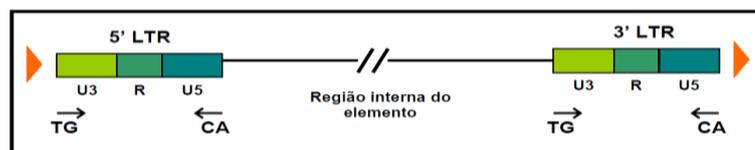


**Figura 3:** Retrotransposons do Tipo LTR, classe II.

**Fonte:** Kalendar et al. (2011)

Uma das principais questões levantadas pelos estudos de evolução dos elementos de transposição é a relação dos retroelementos, como o *gypsy*, que possuem gene *env* potencialmente funcional. Os retrotransposons com LTRs, mais especificamente os elementos da família Ty3, possuem similares estruturas dos retrovírus, como as sequências de LTRs, os genes *gag*, *pol*, e em alguns casos a presença do gene *env* incompleto e não funcional (Kim et al., 2004). Porém análises filogenéticas das sequências nucleotídicas da transcriptase reversa e integrase classificam os elementos desta família (também chamados de *gypsy-like*) como retrotransposons com LTR. Esta é a razão pela qual o elemento *gypsy* tem uma posição filogenética intermediária aos retrovírus e aos retroelementos com LTR (Capy et al., 1998).

A *Long Terminal Repeat* (LTR) geralmente apresenta uma variação de tamanho entre 250 a 600 pb e em sua estrutura comportam três domínios distintos: U3, R e U5 (Figura 4). A transcrição do elemento começa no sítio de iniciação da transcrição de 5' LTR, que ocorre entre os domínios U3 e R. O fim da transcrição ocorre na extremidade 3' LTR onde está o sítio de poliadenilação funcional, que é delimitado pelos domínios R e U5. O domínio U3 contém uma seqüência promotora para a RNA polimerase II, presente em ambas as LTRs, mas somente atua como forte promotor na 5' LTR. A seqüência de 3' LTR poderia atuar como um promotor fraco para os genes adjacentes no local da inserção (White et al., 1994).



**Figura 4:** Organização geral das LTRs de um retrotransposon da superfamília *copia*. Os triângulos (laranja) mostram a orientação do elemento (5' – 3'). Os retângulos (verde) ilustram a organização das LTRs 5' e 3' que flanqueiam a região interna do elemento, com seus respectivos domínios U3, R e U5.

**Fonte:** White et al. (1994).

As informações sobre os retrotransposons no genoma de mandioca ainda são escassas, e até o momento com o seqüenciamento genômico da mandioca em progresso, apenas fragmentos de genes pertencentes a esses retroelementos estão disponíveis em banco de dados genômico, impossibilitando a montagem da seqüência completa de um retrotransposon, principalmente com informações sobre as regiões gênicas *gag*, *pol*, e *env* e as seqüências de LTR. Apesar disso, Gbadegesin et al. (2008) conseguiram isolar e caracterizar, em mandioca, famílias de retrotransposons com LTR, através de fragmentos do gene da enzima transcriptase reversa. Foram caracterizadas 59 famílias *Ty1/copia*-like e 26 famílias *Ty3/gypsy*.

#### ***1.4 Marcador baseado em Retrotransposon – IRAP (Inter-Retrotransposon Amplified Polymorphism)***

Devido à diversidade de retrotransposons, o padrão polimórfico gerado por eles pode ser utilizado para o desenvolvimento de novos marcadores moleculares. Diversos sistemas de identificação de locais de inserção de retrotransposons foram desenvolvidos ao longo dos últimos dez anos e, em princípio, todos eles fornecem o mesmo tipo de

informação. Eles podem ser usados para identificar locais de inserção de elementos transponíveis através de amplificação por PCR (Grzebelus, 2006).

O marcador IRAP foi descrito primeiramente por Kalendar et al. (1999). Ao invés de utilizar enzimas de restrição, essa técnica utiliza iniciadores (*primers*) que amplificam regiões entre retrotransposons, se baseia na amplificação de fragmentos de DNA genômico que se situam entre dois locais de inserção do retrotransposon e utiliza um ou dois primers, desenvolvidos a partir da região de LTR, fazendo com que o trecho de DNA entre dois retrotransposons nas proximidades seja amplificado.

Na descrição da metodologia para IRAP verifica-se que a amplificação pode ser realizada com um único *primer* correspondente a região 5' ou 3' da LTR, mas que é orientado para fora da LTR em si, ou com dois *primers*. Os dois *primers* podem ser da mesma família de elementos ou de famílias diferentes e os produtos da PCR resultam na amplificação de centenas ou milhares de sítios-alvo no genoma.

Os retrotransposons próximos podem ser encontrados em diferentes orientações no genoma (cabeça-cabeça, cauda-cauda, ou cabeça-cauda) aumentando o leque de ferramentas disponíveis para detectar polimorfismo (Kalendar, 2011). Os marcadores moleculares com base em retrotransposons são excelentes ferramentas para a detecção de diversidade genética e alterações genômicas associadas esses elementos genéticos.

Em plantas, o método de marcador IRAP têm sido usado em uma variedade de aplicações, incluindo a medição da diversidade genética e estrutura populacional em maçã (Antonius-Klemola et al. 2006), arroz (Branco et al. 2007), tabaco (Yang et al. 2007), linho (Smýkal et al. 2011), *Pinus massoniana* (Fan et al. 2013).

Essa classe de marcador ainda não foi explorada para estudo de diversidade e estruturação genética em mandioca. O polimorfismo gerado pode ser informativo quanto à história evolutiva da espécie, especialmente considerando que esses elementos existem em abundância no genoma de plantas.

Esta dissertação reúne-se dedica a caracterizar o germoplasma de mandioca da Embrapa com 430 acessos, entre brava, mansa e mandiocaba, coletadas ao longo da bacia amazônica nas calhas dos Rios Madeira, Negro, Purus, Solimões e Baixo Amazonas, por meio de marcadores baseados em retrotransposon LTR, para avaliar distribuição da diversidade genética apresentada pelas variedades de mandioca e estimar a estrutura genética presente na coleção, na tentativa de relacionar essa estrutura à origem geográfica e/ou uso da raiz.

## **2. OBJETIVOS**

### **2.1 OBJETIVO GERAL**

Avaliar a diversidade genética da mandioca por meio de marcadores moleculares com base em elementos transponíveis e avaliar a estrutura genética partir da coleção amazônica de germoplasma de mandioca da Embrapa Amazônia Ocidental.

### **2.2 OBJETIVOS ESPECÍFICOS**

- Desenvolver iniciadores específicos para mandioca a partir de sequências de retrotransposons depositadas no banco de dados genômico da mandioca;
- Analisar a diversidade e estrutura genética do germoplasma da coleção de mandioca da Embrapa Amazônia Ocidental utilizando seis marcadores *Inter-Retrotransposon Amplified Polymorphism* (IRAP).

### **3. MATERIAIS E MÉTODOS**

#### **3.1 Material Vegetal**

Foram utilizados 430 genótipos (entre mandioca doce, amarga e mandiocabas) do Banco de Germoplasma (BAG) da Embrapa Amazônia Ocidental, coletados em diferentes localidades da bacia amazônica.



**Figura 5:** Banco de germoplasma de mandioca da Embrapa Amazônia Ocidental.

**Foto:** Sandra Souza, 2013.

Distribuído pelo uso da raiz de mandioca e pela hidrografia o BAG contém 123 amostras coletadas ao longo do rio Solimões (93 amargas, 29 doces e uma mandiocabas), 124 do Baixo Amazonas (96 amargas, 19 doces e nove mandiocabas), 143 do Rio Madeira (115 amargas e 28 doces), 16 do Rio Purus (11 amargas e cinco doces) e 24 do Rio Negro (22 amargas e duas doces).

O DNA total foi extraído pelo método CTAB 2% no Laboratório de Biologia Molecular da Embrapa.

#### **3.2 Desenvolvimento de Marcadores para IRAP e condições de PCR**

As sequências de retrotransposons LTR (*Long Terminal Repeat*) foram localizadas no banco de dados de genoma de Plantas Phytozome (<http://www.phytozome.net/>) as LTRs foram confirmadas através do software LTR\_Finder ([http://tlife.fudan.edu.cn/ltr\\_finder/](http://tlife.fudan.edu.cn/ltr_finder/)) e

analisadas utilizando BLAST no GenBank (<http://www.ncbi.nlm.nih.gov/BLAST/>). Dos retrotransposons LTR selecionados foram desenhados primers para IRAP a partir da sequência de LTR e projetados para o exterior das LTRs nas orientações como cabeça-cabeça, cabeça-cauda e cauda – cauda, através do software Primer3 (Rozen e Skaletsky, 2000).

A reação de PCR seguiu as recomendações de Kalendar et al. (1999) com algumas modificações, foi otimizada com 50ng de DNA, 0,2 mM de dNTPs, 0,5 µM de cada par de primers F e R, 2 mM de MgCl<sub>2</sub>, solução tampão contendo 50mM de KCl, 10mM de Tris-HCl, pH 8,9, 1,5 U de Taq GoTaq® DNA-polimerase (Promega, USA) em um volume total de 20 µL. O ciclo total de amplificação foi realizado em termociclador Veriti (Applied Biosystems, USA) programado para iniciar com 2 minutos a 92°C seguidos por 40 ciclos com 15 segundos de desnaturação a 92°C, 1 minuto de anelamento com temperaturas variando entre 40-60°C (Tabela 1) e 2 minutos de extensão a 72°C com 10 minutos a 72°C para extensão final.

Os produtos de amplificação foram separados, sob condição desnaturante, em gel de agarose 1,5% em sistema de eletroforese contendo tampão TBE 0,5X com brometo de etídio e digitalizados em fotodocumentador (Loccus Biotecnologia, BRA). O polimorfismo gerado pelos marcadores IRAP foi detectado pela presença (1) ou ausência (0) do produto de PCR.

### ***3.3 Análise dos Dados***

#### ***3.3.1 Estrutura Genética das Populações***

A estrutura genética das populações em *M. esculenta* foi analisada para o conjunto de dados de marcadores IRAP, por meio de análise Bayesiana implementada no software Structure v.2.2 (Pritchard et al., 2000).

Essa análise considera a separação do número total de indivíduos em agrupamentos (clusters), atribuindo-os um número K de populações e assumindo que está em equilíbrio de Hardy-Weinberg. Dessa forma, foram definidas as populações às quais pertenciam os indivíduos, não sendo utilizadas as informações sobre localidade ou qualidade da raiz. O modelo adotado para esta análise foi o admixture, onde cada amostra pode ter ancestrais de mais de uma população.

Vinte simulações foram realizadas utilizando o modelo admixture e as frequências alélicas correlacionadas para cada agrupamento hipotético K das variedades (K variando de 1

a 10), com 500.000 de interações “Monte-Carlo Markov Chains (MCMC)” e *burn-in* de 100.000 para cada simulação. Foram calculados  $\Delta K$  para cada valor de  $K$  (Evanno et al., 2005), o que permite a fácil interpretação de qual  $K$  é mais provável para representar o número de grupos na matriz de dados. Se uma variedade tiver valor maior que o limite arbitrário de 80% do seu genoma em um agrupamento então pertence a esse agrupamento, de acordo com Vigouroux et al. (2008).

### ***3.3.2 Diversidade Genética com marcadores IRAP***

Os índices de diversidade genética, tais como, percentual de loci polimórficos, heterozigosidade esperada ( $h$  = diversidade genética de Nei), Índice de informação de Shannon ( $I$ ) e coeficiente de diferenciação gênica ( $F_{ST}$ ), foram estimados para cada grupo de variedades de  $K = 2$ ,  $K = 3$  e  $K = 5$  através do software Popgen 1.31 (Yeh et al. 1999),

Para examinar a distribuição hierárquica de diversidade genética dos grupos  $K = 2$ ,  $K = 3$  e  $K = 5$  foi avaliado a Análise de Variância Molecular (AMOVA) dentro e entre os grupos através do software GenAlex 6.1 (Peakall and Smouse, 2006).

## 4. RESULTADOS

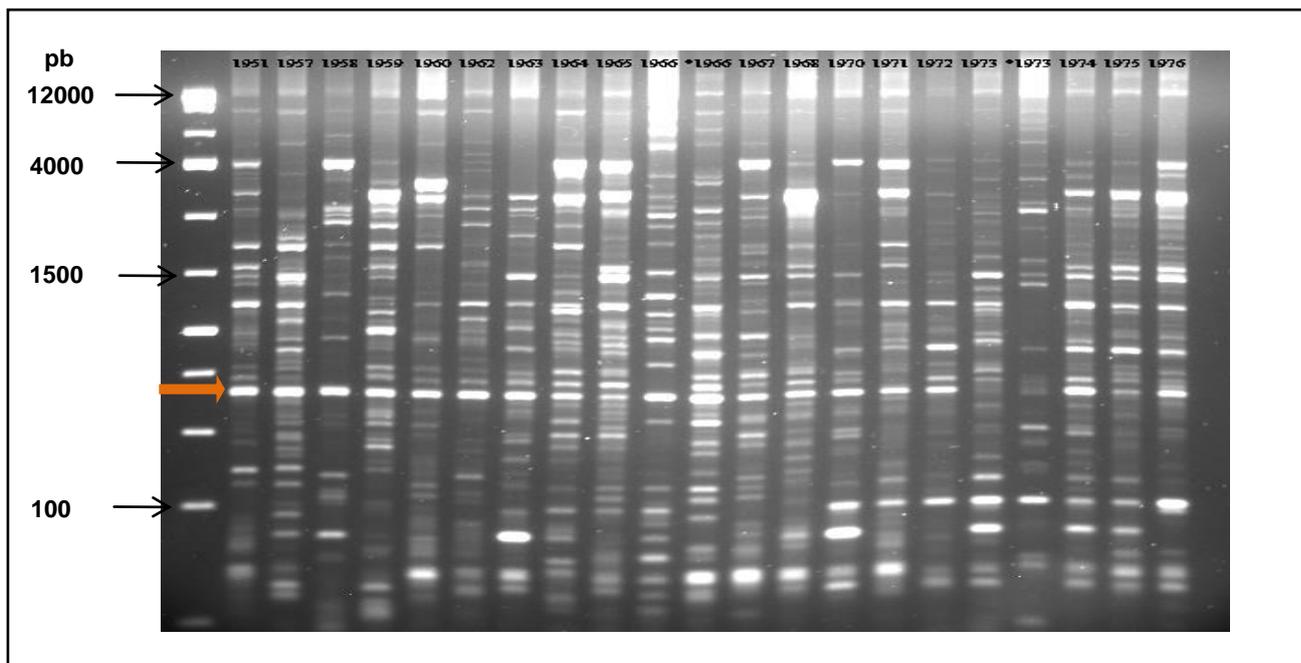
### *Desenho de marcadores para IRAP*

Foram desenvolvidos oito pares de primers IRAP denominados de ME\_1 a ME\_8. Dois dos pares de primers (ME\_1 e ME\_3) apresentaram poucas e fracas bandas e devido ao baixo número de amplicons foram excluídos da análise. Os outros geraram de 23 a 61 bandas com tamanho entre 100 e 12000 bp (Figura 6) e foram publicados na revista Genetic and Molecular Resouce (GMR) (Anexo).

**Tabela 1.** Características de primers IRAP para *Manihot esculenta*.

Nome	Sequencia (5' - 3')	Tamanho (nt)	$T_a$ (°C)	Amplicon (bp)	Scaffold localização Phytozome v.9.0
ME_1	CTGCATTTGAAGTTTGGTCCA <sup>a</sup>	21	60	150 - 1500	00579:8924..18923
	TTTCCAGCTTATTGCTTGGG <sup>b</sup>	20			
ME_2	GGTGATGATGTGTCCCTTCC <sup>a</sup>	20	49	200 - 4000	06700:127431..132763
	CTAGTGATAACCCAATATGCC <sup>b</sup>	22			
ME_3	TCCATCAAATGGGTCTCTCA <sup>a</sup>	20	50	200 - 2000	03122:396999..406998
	ACCCAGCATTTC AAGTCTCG <sup>b</sup>	20			
ME_4	TGGAGCTTGAGGGTGTAAGG <sup>a</sup>	20	48	100 - 5000	00077:42599..52000
	TTCGATTGCTTCTCTCCTGC <sup>b</sup>	20			
ME_5	GCAAGGAGGGGAATAAAAAG <sup>a</sup>	20	40	100 - 4000	03413:6000..16000
	GCTTCTTTCTTACCGGGCTT <sup>b</sup>	20			
ME_6	TTTTTCATTTTCTTACTTTCTGTGTAA <sup>a</sup>	27	42	150 - 1800	09428:1..12686
	CCTATGATTATGCTATCAATTATCAC <sup>b</sup>	26			
ME_7	TTTCTTGATTCCCAAGGGTG <sup>a</sup>	20	48	200 - 3000	01259:102333..122332
	CCTCTCCATATTCTTCTTCTCC <sup>b</sup>	21			
ME_8	GGTGAATTTTCGGTTATTGAA <sup>a</sup>	20	56	100 - 12000	03481:28000..36000
	CCAGAGAATGATGTTGAAGG <sup>b</sup>	20			

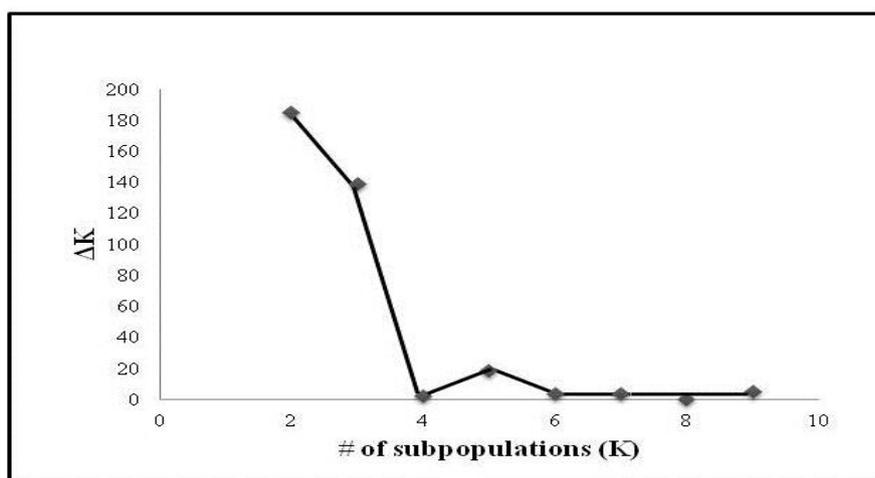
<sup>a</sup> primer desenhado da extremidade LTR 5'; <sup>b</sup> primer desenhado da extremidade LTR 3';  $T_a$  = temperatura de anelamento; bp = pares de base; nt = nucleotídeo.



**Figura 6:** Perfil de amplificação e polimorfismo observado com primer IRAP ME\_8 em variedades do BAG de germoplasma de mandioca. pb = pares de bases.

### *Estrutura Genética em mandioca com marcadores IRAP*

A ferramenta Bayesiana implementada no software STRUCTURE v.2.2 para dados de IRAP (Oliveira-Silva et al., 2014) sugeriu a existência de dois agrupamentos genéticos na coleção de germoplasma de acordo com Evanno et al. (2005), o valor de  $\Delta K$  considerado foi obtido selecionando o melhor número de subpopulações (K) (Figura 7).

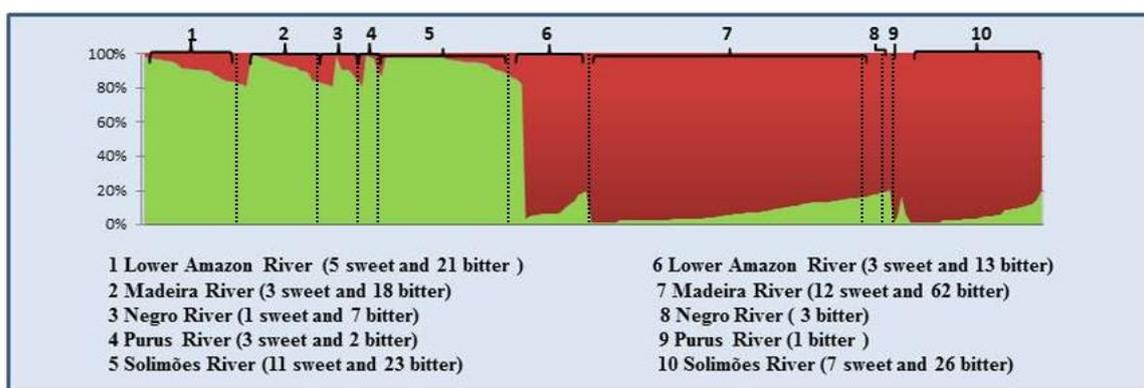


**Figura 7.** Valor de  $\Delta K$  de possíveis agrupamentos de 430 variedades de mandioca brasileira obtidos de dez simulações de análise da estrutura.

Foram analisadas 430 variedades de mandioca, destas somente 220 ( $K = 2$ ) apresentaram fidelidade acima de 80% e essas foram, portanto, utilizadas para análises de diversidade seguindo as recomendações de Vigouroux et al. (2008), o restante das variedades de mandioca (210 amostras) foram consideradas como tendo mistura de ascendência por apresentar valores abaixo de 80% do limite de fidelidade considerado aqui. Além de  $K = 2$  foi sugerido  $K = 3$  razoavelmente significativo e  $K = 5$  insignificante (Figura 7).

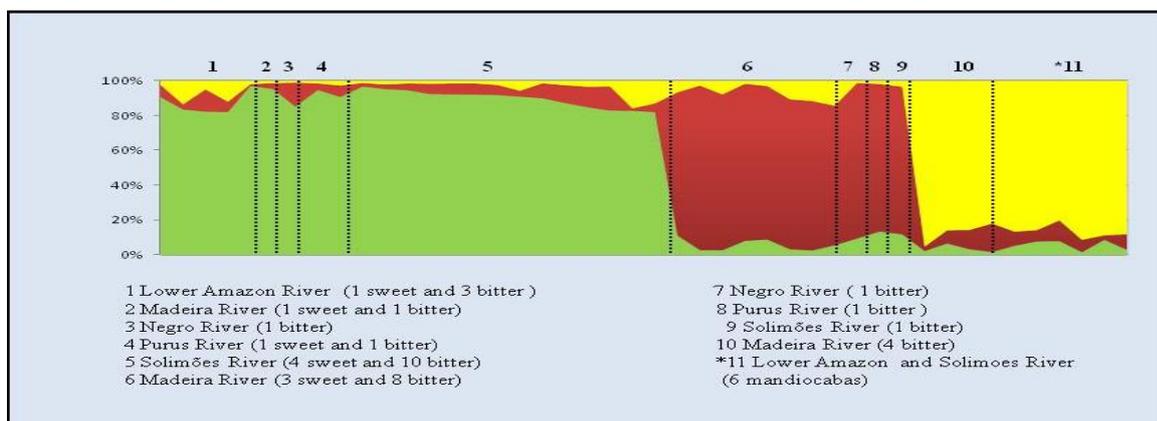
Apenas dois grupos distintos de acessos de mandioca foram observados (Figura 8). Foi confirmado que 93 acessos (23 doces e 70 amargas) pertenciam ao grupo K1, enquanto os outros 127 acessos (22 doces e 105 amargas) foram alocados ao grupo K2 com fidelidade acima de 80%. A Figura 8 mostra os acessos alocados nos grupos K1 e K2.

A AMOVA com base nos dados moleculares mostrou que os grupos obtidos com a análise de estrutura (K1 e K2) diferiram significativamente um do outro, com 89% do restante da variação dentro dos grupos (Tabela 2). O índice de diferenciação entre os grupos foi baixo ( $F_{ST} = 0,043$ ). A análise de variância molecular (AMOVA) para toda a coleção (430 genótipos) em  $K = 2$  apresentou menor variação entre populações (5%) e maior variação dentro das populações (95%), a diferenciação entre os clusters foi menor ainda ( $F_{ST} = 0,027$ ), quando comparada com o agrupamento de 220 genótipos com fidelidade acima de 80%.



**Figura 8.** Agrupamento dos 220 acessos de mandioca de acordo com um algoritmo baseado no modelo Bayesiano implementado no software Structure. Membros das Populações (expressa em%) para o melhor parentesco ( $K = 2$ ).

Em  $K = 3$  apenas 44 variedades apresentaram mais de 80% de fidelidade o restante foi considerado como mistura de ascendência. Existem dois grupos mistos de mandioca amarga e doce, o grupo K1 com mais variedades do Rio Solimões e outros locais, o grupo K2 com mais variedades do Rio Madeira e outros locais, e o grupo K3 constituído por mandioca amarga e mandiocaba (Figura 9) com pelo menos uma ou duas variedades que pertencem a uma das cinco regiões.



**Figura 9.** Agrupamento dos acessos de mandioca de acordo com um algoritmo baseado em modelo Bayesian implementado no programa Structure para (K = 3) com fidelidade acima de 80%.

**Tabela 2.** Média da análise de variância molecular (AMOVA) para agrupamentos por marcadores IRAP para grupos de variedades de mandioca da Bacia Amazônica estimadas no software STRUCTURE. Nível 1 - k = 2 agrupamento acima e abaixo de 80% de fidelidade. Nível 2 - K = 3 agrupamentos. Nível 3 - K = 5.

	Varição	df	Sum of squares	Est. Var.	Varição (%)	Estatística F
<b>K = 2</b> <b>&gt;80%</b>	Entre grupos	1	486,023	4,2	11	<b>F<sub>ST</sub></b> = 0,038
	Dentro dos grupos	218	7658,936	35,1	89	
<b>K = 2</b> <b>&lt;80%</b>	Entre grupos	1	424,677	1,8	5	<b>F<sub>ST</sub></b> = 0,027
	Dentro dos grupos	428	15451,367	36,1	95	
<b>K = 3</b> <b>&gt;80%</b>	Entre grupos	2	359,529	11,15	27	<b>F<sub>ST</sub></b> = 0,186
	Dentro dos grupos	41	1208,925	29,48	73	
<b>K = 3</b> <b>&lt;80%</b>	Entre grupos	2	633,297	2,08	6	<b>F<sub>ST</sub></b> = 0,045
	Dentro dos grupos	427	15242,747	35,697	94	
<b>K = 5</b> <b>&lt;80%</b>	Entre grupos	4	196,703	0,179	0	<b>F<sub>ST</sub></b> = 0,019
	Dentro dos grupos	425	15679,341	36,893	100	

A AMOVA do grupo K = 3 apresentou maior variação entre os grupos e alta diferenciação genética  $F_{ST}$  de 0,186 para variedades com fidelidade acima de 80% e  $F_{ST}$  de 0,045 para variedades com ascendência mista, essa diferenciação entre os agrupamentos pode ter sido devido ao grupo de mandiocabas, porque também não havia estruturação por hidrografia ou o uso da raiz.

Em K = 5, que foi o agrupamento menos significativo, apenas 20 variedades apresentaram mais que 80% de fidelidade, por isso não foi possível calcular AMOVA porque alguns grupos foram representados por apenas uma variedade. Em K = 5 abaixo do limite de

fidelidade de 80%, nenhuma variação foi encontrada entre os grupos apresentando 100% de variação dentro dos grupos, claramente com a menor variação entre as populações ( $F_{ST} = 0,019$ ).

### ***Diversidade Genética dentro do germoplasma de mandioca***

Os 430 acessos foram avaliados por seis pares de marcadores IRAP com total de 205 bandas. A diversidade genética foi calculada para cada grupo através da abordagem Bayesiana implementada no software STRUCTURE v.2.2 para  $K = 2$ ,  $K = 3$  e  $K = 5$ .

Em geral, o percentual de loci polimórficos em nível de agrupamento foi alto, variando de 73% a 98% entre grupos com o limite de fidelidade de 80% (Tabela 3). Assumindo equilíbrio de Hardy-Weinberg, dentro dos clusters para  $K = 2$ , não houve diferença nas estimativas de diversidade considerando ou não o limite de fidelidade de 80%, as variedades de mandioca apresentaram polimorfismo médio de 96% (intervalo de 95,6% a 97%), 0,4 heterozigosidade esperada (0,38-0,42) e 0,57 de Índice de informação de Shannon (0,56-0,6).

Em  $K = 3$  com o limite fidelidade acima de 80%, a média de polimorfismo foi de 77% (variou de 73% a 82%), heterozigosidade esperada de 0,3 (0,27 a 0,34) e Índice de informação de Shannon de 0,44 (0,4 a 0,5) as estatísticas de diversidade não foram afetadas pelo tamanho da amostra, apesar de ter sido avaliada em poucas plantas.

Em  $K = 5$ , embora insignificante, o polimorfismo médio foi de 94% (variou de 92% a 98%), 0,4 de heterozigosidade esperada (0,38 a 0,41) e índice de Informação de Shannon de 0,58 (0,56 a 0,6) entre os clusters, foram bastante semelhantes aos encontrados para  $K = 2$ .

**Tabela 3.** Níveis de polimorfismo e diversidade genética detectados com marcadores IRAP entre mandiocas usadas nesse estudo;  $K=2$  são as melhores subpopulações estimadas na análise do STRUCTURE. \*Número de genótipos, sem considerar genótipos miscigenados entre conjunto de genes. Percentual de loci polimórficos calculado sobre o total de loci (205).

	<b>Grupos</b>	<b>N</b>	<b>Ne</b>	<b>%</b>	<b>*I ± SD</b>	<b>*h ± SD</b>
<b>K = 2 &gt;80%</b>	Grupo 1	93	1,77±0,017	95,6	0,6 ±0,01	0,42 ±0,008
	Grupo 2	127	1,69±0,020	96,5	0,56 ±0,012	0,38 ±0,010
	<b>Total</b>	<b>220</b>	<b>1,7±0,014</b>	<b>95,8</b>	<b>0,57 ±0,008</b>	<b>0,4 ±0,006</b>
<b>K = 2 &lt;80%</b>	*Grupo 1	208	1,77±0,017	97	0,6±0,009	0,42±0,007
	*Grupo 2	222	1,70±0,019	97	0,57±0,011	0,39±0,009

	<b>Total</b>	<b>430</b>	<b>1,74±0,013</b>	<b>97</b>	<b>0,58 ±0,007</b>	<b>0,4 ±0,006</b>
<b>K = 3 &gt;80%</b>	Grupo 1	23	1,6±0,024	82	0,5 ±0,017	0,34 ±0,012
	Grupo 2	11	1,46±0,025	76,6	0,4 ±0,018	0,27 ±0,013
	Grupo 3	10	1,53±0,027	73	0,44 ±0,020	0,30 ±0,014
	<b>Total</b>	<b>44</b>	<b>1,53±0,015</b>	<b>77</b>	<b>0,44±0,011</b>	<b>0,30±0,008</b>
<b>K = 3 &lt;80%</b>	*Grupo 1	110	1,76±0,018	95,6	0,6 ±0,011	0,42 ±0,008
	*Grupo 2	107	1,63±0,020	95,1	0,54 ±0,012	0,36 ±0,009
	*Grupo 3	213	1,75±0,018	97	0,6 ±0,010	0,41 ±0,008
	<b>Total</b>	<b>430</b>	<b>1,7±0,011</b>	<b>96</b>	<b>0,57±0,006</b>	<b>0,4±0,005</b>
<b>K = 5 &lt;80%</b>	*Grupo 1	50	1,93±0,018	93	0,56 ±0,012	0,38 ±0,009
	*Grupo 2	73	1,95±0,015	95	0,59 ±0,011	0,40 ±0,008
	*Grupo 3	44	1,95±0,015	97	0,59 ±0,011	0,41 ±0,008
	*Grupo 4	25	1,92±0,019	92	0,57 ±0,013	0,40 ±0,009
	*Grupo 5	238	1,98±0,010	98	0,6 ±0,009	0,41 ±0,007
	<b>Total</b>	<b>430</b>	<b>1,73±0,008</b>	<b>94</b>	<b>0,58±0,005</b>	<b>0,4±0,004</b>

**Ne** = número de alelos efetivos; **\*h** = diversidade genética de Nei ± desvio padrão; **\*I** = Índice de Informação de Shannon ± SD; percentual de loci polimórfico.

## 5. DISCUSSÃO

Apesar de ser propagada vegetativamente a mandioca tem altos níveis de diversidade genética, como resultado da facilidade de polinização cruzada da espécie, sua alta heterozigosidade e deiscência abrupta dos frutos e devido à maioria das variedades nativas serem selecionadas, muitas vezes pelos próprios agricultores (Fukuda et al., 2006). A diversidade genética dentro e entre os grupos de mandioca foi confirmada principalmente através do uso de marcadores moleculares, incluindo polimorfismo de DNA amplificado ao acaso (RAPD), polimorfismo de comprimento de fragmentos amplificados (AFLP) e sequência simples repetida (SSR) (Colombo et al., 2000; Mühlen et al., 2000; Empeaire et al., 2003).

O percentual de polimorfismo detectado com marcadores IRAP foi alto, com uma média acima de 95% dentro dos clusters estimados pelo STRUCTURE, a heterozigosidade média (0,4) entre os grupos foi semelhante à média de SSRs em mandioca encontradas na região amazônica (Elias et al., 2004) e outras regiões (Montero-Rojas et al., 2011; Turyagyenda et al., 2012; Costa et al., 2013). A média de diversidade com o índice de Shannon foi superior em comparação com médias para marcadores IRAP (Smýkal et al., 2011).

O marcador IRAP não detectou estrutura genética no germoplasma de mandioca para as regiões da Bacia Amazônica ou para o uso da raiz (mandioca amarga e doce) essa falta de estrutura entre as cultivares de mandioca pode ser explicada pela distribuição das inserções de retrotransposons do genoma da mandioca, pois enquanto os retrotransposons são suficientemente estáveis para permitir seu uso em mapeamento (Huo et al., 2009), eles são sensíveis o suficiente para detectar mudanças rápidas no genoma (Belyayev et al., 2010; Petit et al., 2010).

Mühlen et al. (2012) utilizando marcadores microssatélites mostraram agrupamentos separando as variedades de mandioca doce e mandioca amarga semelhantes aos resultados de Empeaire et al. (2003) e Elias et al. (2004) que detectaram estruturação geográfica em variedades de mandioca.

Smýkal et al. (2011) não encontrou nenhuma estrutura no gênero *Linum* através de marcadores IRAP, não houve estrutura genética correspondente ao tipo de uso nem à origem geográfica, muito semelhante aos resultados encontrados aqui para IRAP.

## CONCLUSÕES GERAIS

Este foi o primeiro trabalho com marcadores IRAP para caracterizar as variedades de mandioca. Eles provaram ser bastante informativos ao nível de diversidade genética, eram esperados índices de heterozigosidade.

Existe diversidade genética entre as variedades de mandioca da coleção de germoplasma da Embrapa Amazônia Ocidental e essa diversidade está mais representada dentro dos grupos do que entre eles, no entanto, as análises não suportam a estruturação por origens geográficas nem pelo tipo de uso de variedades de mandioca.

Os marcadores IRAP utilizados não detectaram estruturação dentro do germoplasma de mandioca comparada com a estruturação assistida por outros marcadores, como SSRs e RAPD na região Amazônica.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Allem, A.C. The origins and taxonomy of cassava. In: Hillocks, R. J.; Thresh, J. M.; Bellotti, A. Cassava: biology, production and utilization. New York: Wallingford, UK, p. 1-16, 2002.
- Alves, A.A.C. Cassava botany and Physiology. In: hillocks, R. J.; Thresh, J. M.; Bellotti, A. Cassava: biology, production and utilization. New York: Wallingford, UK, p. 67-89. 2002.
- Antonius-Klemola, K., Kalendar, R., Schulman, A.H. TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports. *Theor. Appl. Genet.* 112:999-1008, 2006.
- Balat M., Balat, H. Recent trends in global production and utilization of bio-ethanol fuel. *Appl Energ.* 86:2273–2282, 2009.
- Barcaccia, G., Alberini, E., Rosellini, D. Inheritance and mapping of 2n-egg production in diploid alfalfa. *Genome.* 43:528-537, 2000.
- Baucom, R. S., J. C. Estill, J. Leebens-Mack, and J. L. Bennetzen. Natural selection on gene function drives the evolution of LTR retrotransposon families in the rice genome. *Genome Res.* 19: 243-254, 2009.
- Belyayev, A., Kalendar, R., Brodsky, L., Nevo, E., Schulman, A.H., Raskina, O. Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat. *Mob DNA.* 1:1-16, 2010.
- Botstein, D., White, R.L., Skolnick, M., Davis, R.V. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314-331, 1980.
- Branco, C.J., Vieira, E.A., Malone, G., Kopp, M.M., Malone, E., Bernardes, A., Mistura, C.C., Carvalho, F.I., Oliveira, C.A. IRAP and REMAP assessments of genetic similarity in rice. *J. Appl. Genet.* 48: 107-113, 2007.
- Capy, P., Anxolabehere, D. e Langin, T. The strange phylogenies of transposable elements: are horizontal transfers the only explanation? *Trends Genet.* 10:7-12, 1994.
- Capy, P., Bazin, C., Higuete, D. e Langin, T. Dynamics and evolution of transposable elements. Austin, Texas: Landes Bioscience. p.197, 1998.
- Chavarriga-Aguirre, P. e Halsey, M.E. Cassava (*Manihot esculenta* Crantz): Reproductive biology and practices for confinement of experimental field trials. Washington, D.C.: Program for Biosafety Systems. 2005.

- Clement, C.R. 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. *Economic Botany*. 53:188-202, 1999.
- Clement, C.R. Melhoria de espécies nativas. In: Nass, L. L.; Valois, A. C. C.; Melo, I. S. de; Valadaresinglis, M. C. (Ed.). Recursos genéticos e melhoria: plantas. Rondonópolis: Fundação MT. p. 423-441, 2001.
- Colombo, C., Second, G., Charrier, A. Diversity within American cassava germ plasm based on RAPD markers. *Genetics and Molecular Biology*. 23: 189-199, 2000.
- Costa, T.R., Filho, P.S.V., Gonçalves-Vidigal, M.C., Galván, M.Z., Lacanallo, G.F., Silva, L.I. Kvitschal, M.V. Genetic diversity and population structure of sweet cassava using simple sequence repeat (SSR) molecular markers. *African Journal of Biotechnology*. 10:1040-1048, 2013.
- Elias, M., Müllen, G.S., McKey, D., Roa, A.C., Tohme, J. Genetic diversity of traditional South American landraces of cassava (*Manihot esculenta* Crantz): an analysis using microsatellites. *Econ. Bot.* 58:242-256, 2004.
- Empereire, L., Mühlen, G.S., Fleury, M.. Approche comparative de la diversité génétique et de la diversité morphologique des maniocs em Amazonie (Brésil et Guyanes). *Les Actes du Bureau des Ressources Génétiques*. 4: 247-267, 2003.
- Empereire, L., Peroni, N. Traditional management of agrobiodiversity in Brazil: a case study of manioc. *Hum. Ecol.* 35:761-768, 2007.
- Evanno, G., Regnaut, S., Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*. 14:2611-2620, 2005.
- Fan, F., Cui, B., Zhang, T., Ding, G., Wen, X. LTR-retrotransposon activation, IRAP marker development and its potential in genetic diversity assessment of masson pine (*Pinus massoniana*). *Tree Genetics & Genomes*. 2013.
- Faraldo, M.I.F., Silva, R.M., Ando, A., Martins, P.S. Variabilidade Genética de Etnovarietades de Mandioca em regiões geográficas do Brasil. *Scientia Agricola*, 57 (3), p.499-505, 2000.
- Ferreira, M.E.; Grattapaglia, D. Introdução ao uso de marcadores moleculares em análise genética. Brasília: Embrapa Cenargen, 220p. 1998.
- Feschotte C. M. A New Superfamily of DNA Transposons Identified in Diverse Animal Genomes and Related to Bacterial IS1016 Insertion Sequences. *Molecular Biology and Evolution*, 21:1769-1780, 2004.
- Fukuda, W.M.G., Costa, I.R.S., Vilarinhos, A.D., Oliveira, R.P. de. Banco de Germoplasma de Mandioca: Manejo, Conservação e Caracterização. Cruz das Almas: EMBRAPA – CNPMF, 103p. 1996.

- Fukuda, W.M.G., Fukuda, C., Dias, M.C., Xavier, J.J.B.N., Fialho, J.F. Variedades. In: Souza LS, Farias ARN, Mattos PLP, Fukuda WMG. Aspectos socioeconômicos e agronômicos da mandioca. Cruz das Almas: Embrapa Mandioca e Fruticultura Tropical. p. 433-454, 2006.
- Gbadegesin, M.A, Wills, M.A and Beeching, J.R. Diversity of LTR-retrotransposons and Enhancer/Suppressor Mutator-like transposons in cassava (*Manihot esculenta* Crantz). *Mol. Genet. Genomics*. 280: 305–317, 2008.
- Grzebelus, D. Transposon insertion polymorphism as a new source of Molecular Markers. *Journal of Fruit and Ornamental Plant Research*. 14(Suppl1), p.9, 2006.
- Hansen, C.N., Heslop-Harrison, J.S. Sequences and Phylogenies of Plant Pararetroviruses, Viruses and Transposable Elements. *Advances in Botanical Research*. 41:165-193, 2004.
- Harlan, J.R. *Crops and man*, 2a. Ed. Madison, WI: American Society of Agronomy and Crop Science Society of America, 284p. 1992.
- Heat, D.D., Iwama, G.K., Devlin, R.H. PCR primed with VNTR core sequences yields species specific patterns and hypervariable probes. *Nucleic Acids Res*. 21:5782-5785, 1993.
- Hershey, C.H. Cassava breeding. CIAT Headgunters. In: Howeler, R.H.; Kawano, K. Cassava breeding and agronomy research in Ásia .Proceedings of a workshop held in Tailand, 1987. Cali, Colombia: CIAT. p. 99 – 116, 1988.
- Hershey, C.H. *Manihot esculenta* diversity. In: International Network for Cassava Genetic Resouces. Cali, Proceedings. Rome: IBPGR, p.111-134, 1992.
- Huo, H., Conner, J.A., Ozias-Akins, P. Genetic mapping of the apospory-specific genomic region in *Pennisetum squamulatum* using retrotransposon-based molecular markers. *Theor Appl Genet*. 119:199–212, 2009.
- Instituto Brasileiro de Geografia e Estatística. Indicadores IBGE. Estatística da Produção Agrícola. Disponível em < [www.sidra.ibge.gov.br](http://www.sidra.ibge.gov.br) > consultado em abril de 2013.
- Jeffreys, A.J., Wilson, V., Thein, S.L. Hypervariable 'minisatelite' regions in human DNA. *Nature*. 316:76-79, 1985.
- Kalendar, R., Grob, T., Regina, M., Suoniemi, A., Schulman, A. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet*. 98:704-711, 1999.
- Kalendar, R., Flavell, A.J., Ellis, T.H.N., Sjakste, T., Moisy, C., Schulman, A.H. Analysis of plant diversity with retrotransposon-based molecular markers. *Heredity*. 106:520-530, 2011.

- Kim, F.J., Battini, J.L., Manel, N., Sitbon, M. Emergence of vertebrate retroviruses and envelope capture. *Virology*. 318:183–191, 2004.
- Kumar, A., Bennetzen, J.L. Plant Retrotransposons. *Annual Review of Genetics*. 33:479-532, 1999.
- Litt, M., Luty, J.A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44:388-396, 1989.
- Lewin, B. Retrovirus and retroposons. In: *Genes VIII*. Carlson G. Ed: Pearson Prentice Hall. NJ. p. 493-509, 2004.
- Lopes, R., Lopes, M.T.G., Oliveira, A.V., Camargo, L.E.A., Fungaro, M.H.P. Carneiro, M.S., Vieira, M.L.C. Marcadores Moleculares Dominantes (Rapid e Aflp). *Biotecnologia Ciência & Desenvolvimento*. 29:56-60, 2002.
- Marco, A. e Marín, I. Retrovirus-like elements in plants. *Recent Res. Devel. Plant Sci, Kerala-India*. 3:2-15, 2005.
- Matioli, S.R. Métodos baseados em PCR para análise de polimorfismo de ácidos nucléicos. In: Matioli SR (Ed.). *Biologia molecular e evolução*. Ribeirão Preto: Holos, 202p. 2001.
- McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA*. 36:344–355, 1950.
- McClintock, B. Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol*. 16:197–216, 1956.
- Milach, S.C.K. Marcadores de DNA. Aplicações no melhoramento de plantas. *UFRGS. Biotecnologia Ciência & Desenvolvimento*. 5:14-17, 1998.
- Montero-Rojas, M., Correa, A.M. Siritunga, D. Molecular differentiation and diversity of cassava (*Manihot esculenta*) taken from 162 locations across Puerto Rico and assessed with microsatellite markers. *AoB Plants*. 2011.
- Mühlen, G.S., Martins, P.S., Ando, A. Variabilidade genética de etnovarietades de mandioca, avaliada por marcadores de DNA. *Scientia Agrícola, Piracicaba*. 57:319-328, 2000.
- Mühlen, G., Alves-Pereira, A., Clement, C.R., Valle, T. Genetic diversity and distribution of manioc (*Manihot esculenta*) in Brazil evaluated with microsatellite markers. 2<sup>nd</sup> International Meeting of Amazonian Archaeology, Manaus, AM. 2012.
- Mullis, K.B., Faloona, F.A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*. 155:335–350, 1987.
- Nass, L.L. Recursos Genéticos Vegetais. 1.<sup>a</sup> edição. Brasília: Embrapa Informação Tecnológica. 440p. 2006.

- Oliveira, E. J., Pádua, J. G., Zucchi, M. I., Vencovsky, R., Vieira, M. L. C. Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, Ribeirão Preto. 29:294-307, 2006.
- Oliveira-Silva, A.M., Silva, G.F., Dias, M.C., Clement, C.R., Sousa, N.R. Inter-retrotransposon-amplified polymorphism markers for germplasm characterization in *Manihot esculenta* (Euphorbiaceae). *Genet. Mol. Res.* 13 (2): 3800-3804, 2014.
- Olsen, K.M. e Schaal, B.A. Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. *PNAS USA*. 96:5586-5591, 1999.
- Peakall, R., Smouse, P.E. Genalex 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. 6:288-295, 2006.
- Peroni, Nivaldo , Hanazaki, N. . Current and lost diversity of cultivated varieties, especially Cassava, under swidden cultivation systems in the Brazilian Atlantic Forest. *Agriculture, Ecosystems and Environment*, 92:171-183, 2002.
- Petit, M., Guidat, C., Daniel, J. Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytol.* 186:135–147, 2010.
- Pritchard, J.K., Stephens, M., Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics*. 155:945-959, 2000.
- Ramakrishna, W; Chowdari, K.; Lagu, M.; Gupta, V. and Ranjekar, P. 1995. DNA fingerprinting to detect genetic variation in rice using hypervariable DNA sequences. *Theor. Appl. Genet. Genet.* 90 :1000-1006, 1995.
- Roa, A.C., Maya, M.M., Duque, M.C., Tohme, J., A.C. Allem e Bonierbale, M.W. AFLP analysis of relationships among cassava and other *Manihot* species. *Theor. Appl. Genet.* 95:741-750, 1997.
- Roa A.C., Chavarriaga-Aguirre, P., Duque, M.C., Maya, M.M., Bonierbale, M.W., Iglesias, C. e Tohme, J. Cross-species amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationship. *Am. J. Bot.* 87:1647-1655, 2000.
- Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics methods and protocols: methods in molecular biology*. S. Misener and S.A. Krawetz, editors. Humana Press. Totowa, New Jersey, USA. 365–386, 2000.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. 239:487–491, 1988.
- Silva, R.M., Bandel, G., Faraldo, M.I.F., Martins, P.S. *Biologia Reprodutiva de Etnovariedades de Mandioca*. *Scientia Agrícola*. 58(1), p.101-107, 2001.

- Singh, Y. Molecular approaches to assess genetic divergence in Rice. GEF Bulletin of Biosciences. 2: 41-48, 2011.
- Smýkal, P., Bacová-Kerteszová, N., Kalendar, R., Corander, J., Schulman, A.H., Pavelek, M. Genetic diversity of cultivated flax (*Linum usitatissimum* L.) germplasm assessed by retrotransposon-based markers. Theor Appl Genet. 122:1385–1397, 2011.
- Stockton, T., Sonnante, G. and Gepts, P. 1990. Detection of minisatellite sequences in *Phaseolus vulgaris*. Plant Mol. Bio. Reporter. 10:47-59, 1990.
- Tomich, R.G.P et al. Etnovarietades de mandioca (*Manihot esculenta* Crantz) cultivadas em assentamentos rurais. MS - Dados eletrônicos - Corumbá: Embrapa Pantanal, (Boletim de Pesquisa e Desenvolvimento/ Embrapa Pantanal, ISSN 1981-7215; 78). 27 p. 2008.
- Turyagyenda, L.F., Kizito, E.B., Ferguson, M.E., Baguma, Y., Harvey, J.W., Gibson, P., Wanjala, B.W. Osiru, D.S.O. Genetic diversity among farmer-preferred cassava landraces In uganda. African Crop Science Journal. 20:15 – 30, 2012.
- Vieira, E.A., Fialho, J.F., Silva, M.S., Fukuda W.M.G., Faleiro, F.G. Variabilidade genética do banco de germoplasma de mandioca da Embrapa cerrados acessada por meio de descritores morfológicos. Revista Científica, Jaboticabal. 36:56 – 67, 2008.
- Vigouroux, Y., Glaubitz, J.C., Matsuoka, Y. Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. American Journal of Botany. 95:1240-1253, 2008.
- Vos, P.; Hogers, R.; Bleeker, M.; Reijans, M.; Van De Lee, T.; Hornes, M.; Frijters, A.; Pot, J.; Peleman, J.; Kuiper, M., Zabeau, M. 1995 AFLP : a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414, 1995.
- Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Losick, M. L. R. Biologia molecular do gene. 5.<sup>a</sup> edição. Porto Alegre: Artmed, 760p. Tradução - Luciane Passaglia & Rivo Fischer. 2006.
- Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, L.A. & Tingey, S.V. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535, 1990.
- White, S.E., Habera, L.F., Wessler, S.R. Retrotransposons in the flanking regions of normal plant genes: A role for copia-like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA. 91:11792-11796, 1994.
- Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z.D., Dubcovsky, J., Keller, B. Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat. Plant Cell. 15:1186-1197, 2003.
- Wicker, T, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8:973–982, 2007.

- Yang, B.C., Xiao, B.G., Chen, X.J., Shi, C.H. Assessing the genetic diversity of tobacco germplasm using intersimple sequence repeat and inter-retrotransposon amplification polymorphism markers. *Ann Appl Biol.* 150:393–401, 2007.
- Yeh, F.C., Yang, R., Boyle, T. PopGen v.1.31. Microsoft Windows-based freeware for population genetic analysis: Quick User Guide. University Alberta/Center for International Forestry Research, Canada, 1–28, 1999.

**ANEXO**

<b>Região/ Rio</b>	<b>Município</b>	<b>Mandioca amarga</b>	<b>Mandioca doce</b>
<b>Baixo Amazonas</b>	Barreirinha	16	1
	Boa Vista do Ramos	10	0
	Boca do Acre	1	0
	Faro	9	0
	Humaitá	2	1
	Itacoatiara	3	1
	Itapiranga	11	2
	Juruti	14	6
	Manaquiri	1	0
	Nhamundá	12	1
	Parintins	6	2
	Pres.Figueiredo	1	0
	S. Sebastião do Uatumã	6	1
Terra Santa	3	0	
Urucará	12	4	
<b>TOTAL</b>		<b>107</b>	<b>19</b>
<b>Madeira</b>	Autazes	2	0
	Borba	21	5
	Humaitá	30	10
	Manicoré	29	5
	Nova Olinda	13	3
	Novo Aripuanã	17	5
	Parintins	1	0
<b>TOTAL</b>		<b>113</b>	<b>28</b>
<b>Negro</b>	Alvarães	1	0
	Barcelos	2	0
	Manacapuru	0	1
	Manaus	7	0
	Novo Airão	1	0
	S.I.do R. Negro	11	1
<b>TOTAL</b>		<b>22</b>	<b>2</b>
<b>Purus</b>	Boca do Acre	3	1
	Canutama	4	0
	Lábrea	3	1
	Pauini	1	3
<b>TOTAL</b>		<b>11</b>	<b>5</b>
<b>Solimões</b>	A. do Norte	1	0
	Anamã	8	1
	Anori	30	7
	Atalaia	0	1
	B. Constant	2	4
	Faro	1	0
	Guajará	0	1
	Iranduba	5	1
	Letícia	2	0
	S. P. de Olivença	11	7
	Tabatinga	15	4
	Tefé	6	1
Uarini	12	3	
<b>TOTAL</b>		<b>93</b>	<b>30</b>

Germoplasma de mandioca da Embrapa Amazônia Ocidental.

# Inter-retrotransposon-amplified polymorphism markers for germplasm characterization in *Manihot esculenta* (Euphorbiaceae)

A.M. Oliveira-Silva<sup>1</sup>, G.F. Silva<sup>2</sup>, M.C. Dias<sup>2</sup>, C.R. Clement<sup>3</sup> and N.R. Sousa<sup>2</sup>

<sup>1</sup>Programa de Pós-Graduação em Biotecnologia,

Universidade Federal do Amazonas, Manaus, AM, Brasil

<sup>2</sup>Laboratório de Biologia Molecular, Embrapa Amazônia Ocidental, Manaus, AM, Brasil

<sup>3</sup>Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, Brasil

Corresponding author: N.R. Sousa

E-mail: nelcimar.sousa@embrapa.br

Genet. Mol. Res. 13 (2): 3800-3804 (2014)

Received February 13, 2013

Accepted July 1, 2013

Published May 16, 2014

DOI <http://dx.doi.org/10.4238/2014.May.16.3>

**ABSTRACT.** Manioc, *Manihot esculenta*, is economically important in many tropical and subtropical countries. The genetic variability of the species has not been fully explored, and new information may help expand its use. Molecular markers based on retrotransposons have good potential for analysis of genetic diversity given their abundance in the genome. Eight long terminal repeat retrotransposons were selected for the development of interretrotransposon-amplified polymorphism markers. To test these primers, we analyzed 32 varieties from Anori, 30 from Manicoré and 10 Mandiocabas from the Manioc Germplasm Bank at Embrapa Western Amazonia. The six informative primer pairs yielded 20-60 polymorphic bands, averaging 92% polymorphism (51.7-98.4) and 0.37 heterozygosity (0.17 to 0.40), with a Shannon information index of 0.54 (0.26-0.59). These markers can be used to explore the genetic diversity of manioc.

**Key words:** Inter-retrotransposon-amplified polymorphism; Polymorphism; Genetic diversity; Varietal discrimination

## INTRODUCTION

Manioc (*Manihot esculenta* Crantz) is a perennial shrub in the Euphorbiaceae that is a major carbohydrate crop in tropical and subtropical countries, where it is mainly used for the production of flour, pure starch, fresh consumption, and assorted industrial uses. Although propagated vegetatively, manioc has great genetic variability because sexual reproduction continues, often resulting in polyclonal varieties (Silva et al., 2001). These local varieties, grown mostly by small-holder farmers, represent the genetic resources conserved and used in breeding programs.

Molecular markers are used to increase the discriminatory power of genetic variability analyses among manioc varieties. Although there are numerous published markers, there is still a need for new and more variable genetic markers, given the polyclonal nature of manioc varieties. Markers based on retrotransposons (IRAP - inter-retrotransposon-amplified polymorphism) generate great quantities of information, making them good tools for detecting genomic changes associated with their activity, because they create large and stable insertions in the genome; they are highly reproducible, show abundant polymorphism, and are easily viewed in a single gel (Kalendar et al., 2011). Retrotransposon polymorphisms are detected using marker systems that rely on PCR amplification between long terminal repeat (LTR) ends and some components of flanking genomic DNA. The IRAP products are generated with one or two primers matching either the 5'- or 3'-end of the LTR using outward-facing primers (Kalendar and Schulman, 2006).

Several families of transposable elements have been reported in manioc (Gbadegesin et al., 2008). This study describes the development of eight sets of IRAP primers based on retrotransposons to discriminate between manioc clones and varieties.

## MATERIAL AND METHODS

Sequences of LTR retrotransposons were located in Phytozome (<http://www.phytozome.net/>). LTRs were confirmed with the LTR\_Finder software ([http://tlife.fudan.edu.cn/ltr\\_finder/](http://tlife.fudan.edu.cn/ltr_finder/)) and cross-checked in GenBank using BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>). Eight LTR retrotransposons were selected and the outward-facing primers of the LTRs were designed using Primer3 (Rozen and Skaletsky, 2000) for amplification of the members of a retrotransposon family in head-to-head, head-to-tail and tail-to-tail orientation.

Seventy-two manioc plants (32 from Anori, Amazonas, 30 from Manicoré, Amazonas, and 10 “mandiocabas”, a very sweet variety) maintained in the Manioc Germplasm Bank at Embrapa Western Amazônia, Manaus, Amazonas, were analyzed to test the information content of these IRAP markers.

Total DNA was extracted with 2% CTAB in the Molecular Biology Laboratory at Embrapa. The 20- $\mu$ L reaction mixture consisted of 0.2 mM dNTPs, 0.5  $\mu$ M forward and reverse primers, 2 mM MgCl<sub>2</sub>, 1X Taq buffer, 1.5 U Taq GoTaq®DNA polymerase (Promega, USA) and 50 ng template DNA. PCR was performed using the following program: 2 min at 92°C; 40 cycles of 15 s at 92°C, 1 min at 40-60°C (Table 1), 2 min at 72°C; and final extension at 72°C for 10 min. Amplifications were carried out in a Veriti Thermal Cycler (Applied Biosystems, USA). PCR products were resolved by electrophoresis on 1.5% agarose gels in 0.5X TBE buffer stained with ethidium bromide and photographed in a transilluminator (Loccus Biotecnologia, Brazil). Polymorphism was detected by the presence (1) or absence (0) of the PCR product. The percentage of polymorphic loci, expected

heterozygosity (Nei's genetic diversity) and Shannon information index were calculated with PopGene 1.31 (Yeh et al., 1999), for each population and overall.

## RESULTS AND DISCUSSION

Two primer pairs (ME\_1 and ME\_3) produced only six bands and were excluded from further analysis. The others generated 20-60 polymorphic bands with sizes between 100 and 12,000 bp (Tables 1 and 2). With these primers, the two populations and mandiocaba samples showed a mean polymorphism of 92% (range = 51.7 to 98.4%), 0.37 expected heterozygosity (0.17 to 0.40) and a Shannon information index of 0.54 (0.26 to 0.59) (Table 2). Guo et al. (2006) found 86% polymorphism with IRAP and REMAP primers in persimmon (*Diospyros kaki* Thunb.), very similar to the results found here. In manioc, 69% polymorphism was previously found with AFLP markers and 56% with RAPD markers in estimating the genetic variability of 54 varieties (Mühlen et al., 2000), demonstrating that these IRAP markers are more informative than other dominant markers.

**Table 1.** Characteristics of IRAP primers for *Manihot esculenta*.

Primer name	Sequence (5'- 3')	Length (nt)	$T_a$ (°C)	Amplicon Size (bp)	Scaffold location Phytozome v.9.0
ME_1	CTGCATTTGAAGTTTGGTCCA <sup>a</sup> TTTCCAGCTTATTGCTTGGG <sup>b</sup>	21 20	60	150 - 1500	00579:8924..18923
ME_2	GGTGATGATGTGTCCCTTCCa CTAGTGATACCCAATATGCCC <sup>b</sup>	20 22	49	200 - 4000	06700:127431..132763
ME_3	TCCATCAAATGGGTCTCTCA <sup>a</sup> ACCCAGCATTCAAGTCTCG <sup>b</sup>	20 20	50	200 - 2000	03122:396999..406998
ME_4	TGGAGCTTGAGGGTGTAAAG <sup>a</sup> TTCGATTGCTTCTCTCTGC <sup>b</sup>	20 20	48	100 - 5000	00077:42599..52000
ME_5	GCAAGGAGGGGGAATAAAAAG <sup>a</sup> GCTTCTTCTTACCGGGCTT <sup>b</sup>	20 20	40	100 - 4000	03413:6000..16000
ME_6	TTTTTCATTTTCTTACTTTCTGTGTAA <sup>a</sup> CCTATGATTATGCTATCAATTATCAC <sup>b</sup>	27 26	42	150 - 1800	09428:1..12686
ME_7	TTTCTTGATTCCCAAGGGTG <sup>a</sup> CCTCTCCATATTCTTCTCTCC <sup>b</sup>	20 21	48	200 - 3000	01259:102333..122332
ME_8	GGTGAATTTCCGGTTATTGAA <sup>a</sup> CCAGAGAATGATGTTGAAG <sup>b</sup>	20 20	56	100 - 12000	03481:28000..36000

<sup>a</sup> primer designed from 5' LTR end; <sup>b</sup> primer designed from 3' LTR end;  $T_a$  = annealing temperature; nt = nucleotide.

**Table 2.** Genetic information of each IRAP primer pair in 72 *Manihot esculenta* varieties in the Manioc Germplasm Bank at Embrapa Western Amazonia.

Primer	Samples	No. of plants	No. polymorphic bands	$h \pm SD$	$I \pm SD$	%
ME_2	Anori	32	25	$0.38 \pm 0.17$	$0.55 \pm 0.23$	89.3
	Manicoré	30	24	$0.32 \pm 0.18$	$0.47 \pm 0.24$	85.7
	Mandiocaba	10	19	$0.26 \pm 0.20$	$0.38 \pm 0.28$	67.8
	Total	72	25	$0.36 \pm 0.16$	$0.52 \pm 0.22$	89.2
ME_4	Anori	32	26	$0.30 \pm 0.18$	$0.45 \pm 0.24$	89.6
	Manicoré	30	24	$0.30 \pm 0.19$	$0.44 \pm 0.26$	82.7
	Mandiocaba	10	15	$0.17 \pm 0.19$	$0.26 \pm 0.27$	51.7
	Total	72	26	$0.32 \pm 0.18$	$0.47 \pm 0.26$	89.6
ME_5	Anori	32	33	$0.34 \pm 0.16$	$0.50 \pm 0.22$	89.2
	Manicoré	30	32	$0.31 \pm 0.17$	$0.46 \pm 0.24$	86.5
	Mandiocaba	10	24	$0.25 \pm 0.21$	$0.37 \pm 0.30$	64.8
	Total	72	33	$0.36 \pm 0.15$	$0.52 \pm 0.21$	89.2
ME_6	Anori	32	21	$0.36 \pm 0.17$	$0.52 \pm 0.23$	87.5
	Manicoré	30	22	$0.34 \pm 0.18$	$0.50 \pm 0.24$	91.6
	Mandiocaba	10	18	$0.27 \pm 0.19$	$0.41 \pm 0.27$	75
	Total	72	22	$0.39 \pm 0.17$	$0.55 \pm 0.22$	91.6
ME_7	Anori	32	19	$0.30 \pm 0.17$	$0.45 \pm 0.24$	82.6
	Manicoré	30	19	$0.31 \pm 0.19$	$0.45 \pm 0.26$	82.6
	Mandiocaba	10	13	$0.23 \pm 0.22$	$0.33 \pm 0.31$	56.5
	Total	72	20	$0.32 \pm 0.18$	$0.48 \pm 0.24$	90
ME_8	Anori	32	59	$0.39 \pm 0.12$	$0.57 \pm 0.15$	96.7
	Manicoré	30	60	$0.40 \pm 0.11$	$0.59 \pm 0.14$	98.4
	Mandiocaba	10	52	$0.31 \pm 0.17$	$0.46 \pm 0.23$	85.2
	Total	72	60	$0.42 \pm 0.09$	$0.61 \pm 0.11$	98.4
Mean				$0.37 \pm 0.15$	$0.54 \pm 0.20$	92

$h$  = Nei's genetic diversity  $\pm$  standard deviation (SD);  $I$  = Shannon information index  $\pm$  SD; % = percentage of polymorphic loci.

## CONCLUSIONS

This is the first report of IRAP markers to characterize manioc varieties. They proved to be efficient in estimating the percentage of polymorphism and genetic diversity, and will likely permit good variety discrimination.

## ACKNOWLEDGMENTS

Research supported by EMBRAPA (Project #0106010071406; Manioc Regional Germplasm Bank), the Foundation of the State of Amazonas (FAPEAM; Master's scholarship to A.M. Oliveira-Silva) and the National Research Council (CNPq; Research fellowship to C.R.

Clement). We thank Doriane P. Rodrigues, Federal University of Amazonas, for help with the analysis.

## REFERENCES

- Gbadegesin MA, Wills MA and Beeching JR (2008). Diversity of LTR-retrotransposons and Enhancer/Suppressor Mutator-like transposons in cassava (*Manihot esculenta* Crantz). *Mol. Genet. Genomics* 280: 305-317.
- Guo D, Zhang H and Luo Z (2006). Genetic relationships of *Diospyros kaki* Thunb. and related species revealed by IRAP and REMAP analysis. *Plant Sci.* 170: 528-533.
- Kalendar R and Schulman AH (2006). IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. *Nat. Protoc.* 1: 2478-2484.
- Kalendar R, Flavell AJ, Ellis TH, Sjakste T, et al. (2011). Analysis of plant diversity with retrotransposon-based molecular markers. *Heredity* 106: 520-530.
- Mühlen GS, Martins OS and Ando A (2000). Variabilidade genética de etnovariedades de mandioca, avaliada por marcadores de DNA. *Sci. Agric.* 57: 319-328.
- Rozen S and Skaletsky HJ (2000). PRIMER3 on the WWW for General Users and for Biologist Programmers. In: *Bioinformatics Methods and Protocols Methods in Molecular Biology* (Krawetz S and Misener S, eds.). Humana Press, Totowa, 365-386.
- Silva RM, Bandel G, Faraldo MIF and Martins PS (2001). Biologia reprodutiva de etnovariedades de mandioca. *Sci. Agric.* 58: 101-107.
- Yeh FC, Yang R and Boyle T (1999). PopGen v.1.31. Microsoft Windows-Based Freeware for Population Genetic Analysis: Quick User Guide. University of Alberta/Center for International Forestry Research, Alberta.