**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

Diversidade genética, genômica e filogeografia de mandioca (*Manihot esculenta* Crantz): implicações para a dispersão do cultivo ao longo dos principais eixos fluviais da bacia amazônica brasileira

Alessandro Alves Pereira

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Genética e Melhoramento de Plantas

**Piracicaba
2015**

Alessandro Alves Pereira
Biólogo Bacharel em Genética e Biotecnologia

**Diversidade genética, genômica e filogeografia de mandioca (*Manihot esculenta* Crantz):
implicações para a dispersão do cultivo ao longo dos principais eixos fluviais da bacia
amazônica brasileira**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:
Profa. Dra. **MARIA IMACULADA ZUCCHI**

Tese apresentada para obtenção do título de Doutor em
Ciências. Área de concentração: Genética e Melhoramento
de Plantas

**Piracicaba
2015**

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Pereira, Alessandro Alves

Diversidade genética, genômica e filogeografia de mandioca (*Manihot esculenta* Crantz): implicações para a dispersão do cultivo ao longo dos principais eixos fluviais da bacia amazônica brasileira / Alessandro Alves Pereira. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2015.
138 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. *Manihot esculenta* 2. cpSSR 3. ncSSR 4. SNPs 5. Estrutura genética 6. Genômica de populações 7. Cultivos amazônicos I. Título

CDD 633.4
P436d

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICO,

Aos meus pais Raimundo e Avani por toda a sua vida de dedicação, devoção, investimento e amor dedicados a mim. À minha família, meu lar e porto seguro. Amo todos vocês. Aos meus amigos por todo apoio e companheirismo.

Agradecimentos

À FAPESP pela concessão do auxílio à pesquisa (2013/00003-0) e bolsa de estudos (2013/11137-7) na metade final do período de doutorado, que foram fundamentais para a realização deste trabalho e minha formação;

Ao CNPq e CAPES pela concessão de bolsas de estudo na metade inicial do período de doutorado;

À Dr^a. Maria Imaculada Zucchi, chefe (mandona), mãezona e conselheira. Obrigado pela honra de poder ter sido seu aluno. Para você não basta agradecer somente pela confiança e oportunidade. Obrigado por todos os (muitos) trabalhos que confiou a mim, pelos incentivos, opiniões e piadas. Obrigado pela ajuda na confecção dos projetos de auxílio à pesquisa e bolsa. Obrigado pelas aulas que pediu para eu ministrar. Neste tempo aprendi muito, não ganhei somente experiência e conhecimentos científicos, mas também apreendi valores para a vida toda. Sinto que tudo o que eu escrever não será suficiente, mas mesmo assim muito obrigado;

Ao Dr. José Baldin Pinheiro, pela orientação no início do doutorado. Obrigado por ter aberto as portas do seu grupo de pesquisas, e por não medir esforços para oferecer a todos o melhor. Agradeço a disponibilização do espaço físico para a realização dos experimentos. Obrigado também por achar que eu poderia contribuir com seu grupo;

À Universidade de São Paulo, por meio da Escola Superior de Agricultura “Luiz de Queiroz”, e do Departamento de Genética pela oportunidade de realização do curso de Doutorado e pela estrutura física e recursos humanos;

Ao Dr. Charles R. Clement (INPA) pelo apoio financeiro e intelectual no início deste trabalho. Pelo incentivo e ajuda durante o decorrer do doutorado. Pela amizade, e por ter me dado a oportunidade de iniciar os estudos com mandioca. Incluo nesta última parte do agradecimento também o Dr. Nivaldo Peroni (UFSC);

A todos os agricultores tradicionais nas localidades visitadas, por dedicarem seu tempo e permitirem as coletas de amostras das mandiocas;

Aos companheiros do LEA (ICB-UFAM), Dr^a. Doriane Rodrigues, aos Doutores André Junqueira e Catarina Jakovac, aos Mestres Juliana Lins, Leonardo Kumagai, Michelly Araújo, Priscila Ambrósio, Rubana Palhares, Tiara Cabral, e à Karyne Glück (in memoriam), pela boa convivência e auxílio nas coletas das variedades de mandioca;

À Dr^a. Elizabeth Ann Veasey (ESALQ/USP) e seus alunos e meus amigos Dr. Santiago Ramos e MSc. Gabriel Dequigiovanni pelo apoio financeiro e logístico na coleta de variedades de mandioca e variedades silvestres em Rondônia e no Pará;

À Jaqueline Campos, Aline Moraes e Patrícia Zambon, pela ajuda nos procedimentos de detecção de polimorfismo de marcadores microssatélites;

Ao João Paulo pela amizade e ajuda na computação em Linux e em nuvem;

To Tyler Linderoth for discussions about genomics and next-generation sequence data analysis;

Aos meus amados pais, Raimundo e Avani, e à toda a minha família, por todo amor, e especialmente por todos os sacrifícios que fizeram para me dar a melhor educação e ensino possíveis, e que me fizeram chegar até aqui. Todo meu esforço e trabalho é para vocês;

Aos amigos do LabZucchi-MariaRules (Grupo de Genética e Genômica da Conservação) Camila, Carolina, Éllida, Gustavo, Jaqueline, João, Kaiser, Marcos, Mariana, Patricia, Vitor, pela convivência, discussões e sugestões que de alguma forma me auxiliaram neste trabalho;

Aos amigos e colegas que estão ou passaram pelo Laboratório de Diversidade Genética e Melhoramento, em especial Carlos, Fátima, Fernanda, Kênia, Maisa, Michele e Miklos, pelos momentos dentro e fora do laboratório que também me ajudaram a completar esta etapa;

Aos amigos do café, Gabriel, Glaucia, Melina, Santiago, Sueme, Tatiane e Thais, pelos momentos de descontração. De vocês só vou levar coisas boas porque ajudaram a fazer meu período de doutorado melhor. Obrigado pelo companheirismo, ajudas, amizade e carinho;

Às amigas e amigos Ana Beatriz Zanardo, Fabiano Lucas, Francisca Holanda, Leila Camacho, Maisa Curtolo, Marcones Costa, Natália Leite, Natasha Iwanicki, Ricardo Celestino, Sofia Fogliata e Sueme Ueno, pela confiança em mim depositada. Me sinto responsável por pelo menos um pouco da formação de vocês. Vocês não me deram trabalho, eu que agradeço por ter a oportunidade de ajudar e aprender coisas novas com vocês e com vossas pesquisas;

A todos os professores e alunos do Departamento de Genética da ESALQ/USP, pela transmissão e compartilhamento de conhecimento;

Aos amigos de Manaus que, mesmo longe, estão comigo, e a todos aqueles que de forma direta ou indireta tenham contribuído para a realização deste trabalho;

A Deus, pela vida na qual pude conhecer e conviver com todas essas pessoas.

*“I just came across a manger where there is no danger
where love has eyes and is not blind.”*

(Damien Rice)

*“... you’re owing to yourself and to everyone else to continue to stay curious, and by doing so
you’ll have created the meaning of the life, that others think is waiting under a rock for them
one day.”*

(Neil deGrasse Tyson)

*“If you had one shot, or one opportunity, to seize everything you ever wanted, one moment.
Would you capture it, or just let it slip?”*

(Marshall Bruce Mathers III)

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	21
1 INTRODUÇÃO	23
1.1 Objetivos.....	25
1.1.1 Objetivo geral	25
1.1.2 Objetivos específicos.....	25
1.2 Hipóteses	25
2 REVISÃO BIBLIOGRÁFICA	27
2.1 Aspectos gerais da mandioca.....	27
2.2 Domesticação do cultivo	29
2.3 Divergência entre grupos de variedades cultivadas.....	31
2.4 Dinâmica evolutiva de mandioca sob domesticação	34
2.5 Marcadores moleculares microssatélites	36
2.6 <i>Restriction-site associated DNA sequencing</i> (RAD-seq) e marcadores SNPs.....	38
2.7 Genômica populacional	42
2.8 Filogeografia.....	44
3 MATERIAL E MÉTODOS	47
3.1 Material Vegetal	47
3.2 Extração e quantificação de DNA	50
3.3 Amplificação e detecção do polimorfismo de marcadores cpSSR.....	50
3.4 Amplificação e detecção do polimorfismo de marcadores ncSSR.....	51
3.5 Detecção do polimorfismo de marcadores SNPs com a técnica de RAD-seq.....	52
3.6 Análises Genéticas.....	54
3.6.1 cpSSR	54
3.6.2 ncSSR	56

3.6.3 RAD-seq.....	57
4 RESULTADOS.....	61
4.1 Diversidade e estrutura haplotípica de variedades de mandioca com base em cpSSR.....	61
4.2 Diversidade e estrutura genética de variedades de mandioca com base em ncSSR	68
4.3 Diversidade e estrutura genômica de variedades de mandioca com base em marcadores SNPs.....	79
5 DISCUSSÃO	101
5.1 Diversidades haplotípica, genética e genômica de variedades de mandioca	101
5.2 Padrões de estruturação haplotípica, genética e genômica, e agrupamentos de variedades de mandioca	107
5.3 Marcadores SNPs hipoteticamente sob seleção	112
5.4 Considerações sobre a dispersão do cultivo.....	117
6 CONCLUSÃO	121
REFERÊNCIAS.....	123

RESUMO

Diversidade genética, genômica e filogeografia de mandioca (*Manihot esculenta* Crantz): implicações para a dispersão do cultivo ao longo dos principais eixos fluviais da bacia amazônica brasileira

A mandioca foi domesticada no sudoeste da bacia amazônica, e presentemente é o cultivo alimentício amazônico mais importante no mundo. Após a domesticação inicial pressões seletivas divergentes deram origem aos grupos de variedades de mandiocas mansas e bravas. A distribuição atual destes grupos é um tanto diferente ao longo da Amazônia, o que pode ser reflexo de padrões de dispersão distintos de variedades mansas e bravas ao longo da história da domesticação do cultivo. O objetivo deste estudo foi avaliar a diversidade e estrutura genética, genômica e a filogeografia de mandiocas cultivadas por agricultores tradicionais ao longo dos principais rios da bacia amazônica brasileira. Análises filogenéticas de linhagens matrilineares foi realizada com base no polimorfismo de quatro marcadores microssatélites cloroplastidiais (cpSSR). A diversidade e estrutura genética foram avaliadas com 14 marcadores microssatélites nucleares (ncSSR), enquanto que a abordagem genômica foi realizada com base em 5.871 polimorfismos de nucleotídeos únicos (SNPs). Foi observada considerável diferenciação [F_{ST} = 0,78 (cpSSR), 0,28 (ncSSR), e 0,37 (SNPs)] entre as variedades cultivadas e *Manihot esculenta* ssp. *flabellifolia*, o parente silvestre da mandioca. Não foram detectadas associações de haplótipos cpSSR com os grupos de variedades mansas e bravas ou com os rios. Apesar da ausência de padrões filogeográficos, as análises de agrupamento e estrutura genética com base nos três tipos de marcadores avaliados sugeriram que as variedades mansas e bravas são igualmente relacionadas à população silvestre. O segundo padrão mais importante de estruturação genética foi observado entre variedades mansas e bravas [F_{ST} = 0,08 (ncSSR) e 0,10 (SNPs)], no entanto houve um considerável grau de mistura entre variedades de ambos os grupos. Estes resultados, juntamente com a elevada variação genética observada dentro de variedades mansas e bravas são resultantes do manejo realizado por agricultores tradicionais ao longo da região amazônica. A ausência de padrões filogeográficos entre rios e regiões foi observada também com marcadores ncSSR e SNPs. Entretanto, quando a estrutura genética e genômica foi avaliada dentro das variedades mansas e bravas, alguns padrões contrastantes e tendências de estruturação entre rios foram observados. A ausência de padrões claros de estrutura genética e genômica entre diferentes rios não permitiu inferências sobre prováveis rotas de dispersão do cultivo a partir do seu centro de origem no sudoeste da Amazônia. No entanto, os padrões contrastantes de diferenciação genética e genômica dentro de variedades mansas e bravas podem estar associados a histórias de dispersão distintas para estes grupos de variedades de mandioca. Entre os locos genômicos, 658 SNPs estão possivelmente sob seleção positiva quando se considera a divergência entre variedades de mandioca cultivada e o parente silvestre. Destes, 202 SNPs podem estar especificamente associados com a seleção divergente entre variedades mansas e bravas. Estes locos podem estar em genes importantes para a domesticação inicial e seleção para características importantes do cultivo, e podem ser o ponto de partida para o melhor entendimento das bases genômicas da domesticação e diversificação da mandioca.

Palavras-chave: *Manihot esculenta*; cpSSR; ncSSR; SNPs; Estrutura genética; Genômica de populações; Cultivos amazônicos

ABSTRACT

Genetic diversity, genomics and phylogeography of manioc (*Manihot esculenta* Crantz): Implications for the dispersal of the crop along the main fluvial axes in Brazilian Amazonia

Manioc, or cassava, was domesticated in southwestern Amazonia, and is currently the most important staple crop in the world that originated there. After its initial domestication divergent selective pressures gave rise to the groups of sweet and bitter varieties. The current distribution of these groups is somewhat different across Amazonia, which may be due to distinct dispersal patterns of sweet and bitter varieties during the crop's domestication history. The aim of the present study was to evaluate genetic diversity and structure, genomics, and phylogeography of manioc cultivated by traditional farmers along the major rivers of Brazilian Amazonia. Phylogenetic analyses among matrilineages were performed based on the polymorphism of four chloroplastidial microsatellite markers (cpSSR). The evaluation of genetic diversity and structure were performed with 14 nuclear microsatellite markers (ncSSR), and a genomics approach was performed based on 5,871 single nucleotide polymorphism markers (SNPs). Considerable differentiation [$F_{ST} = 0.78$ (cpSSR), 0.28 (ncSSR), and 0.37 (SNPs)] was observed between cultivated varieties and *Manihot esculenta* ssp. *flabellifolia*, manioc's wild relative. No associations of cpSSR haplotypes with the groups of sweet and bitter varieties, nor with rivers were detected. Despite the lack of phylogeographic patterns, the analyses of genetic structure and relationships suggested that sweet and bitter varieties are equally related to wild populations. The second most important pattern of genetic structuring was observed between sweet and bitter varieties [$F_{ST} = 0.08$ (ncSSR) and 0.10 (SNPs)], although there was considerable overlap between groups. These results, combined with the high levels of genetic variability observed within sweet and bitter varieties, are due to the traditional management practices of smallholder farmers across Amazonia. The lack of phylogeographical patterns among rivers and regions were also observed with ncSSR and SNP markers. However, when the genetic and genomic structures were separately evaluated within sweet and bitter varieties, some contrasting patterns and tendencies of genetic structuring among the rivers was observed. The absence of clear patterns of genetic and genomic structure among different rivers did not permit inferences on probable routes of dispersal of the crop from its center of origin in southwestern Amazonia. Nevertheless, the contrasting patterns of genetic and genomic differentiation within sweet and bitter varieties may be associated with distinct dispersal histories for these groups of manioc varieties. Among the genomic loci, 658 SNPs are possibly under positive selection when considering the divergence between cultivated varieties of manioc and the wild relative. Of these, 202 SNPs may be specifically associated with divergent selection between sweet and bitter varieties. These loci may be located in genes important for initial domestication and selection for important characteristics of the crop, and may be a starting point for better comprehension of the genomic bases of manioc domestication and diversification.

Keywords: *Manihot esculenta*; cpSSR; ncSSR; SNPs; Genetic structure; Population genomics; Amazonian crops

LISTA DE FIGURAS

- Figura 1 - Aspectos morfológicos da planta de mandioca (*Manihot esculenta* ssp. *esculenta*). Fotografias ilustrando A) hábito arbustivo de uma variedade de mandioca brava; B) aspecto das raízes logo após a colheita; C) raízes descascadas antes do processamento de detoxificação (Fotografias por André Junqueira)27
- Figura 2 - Variação morfológica presente em variedades cultivadas representada por: variedades de mandioca brava cultivada em A) solos antropogênicos; B) Latossolos; e variedade mansa cultivada em C) solo de várzea. São perceptíveis variações na cor e formato das folhas e na cor dos pecíolos. (Fotografias por André Junqueira)28
- Figura 3 - *Manihot esculenta* ssp. *flabellifolia*. Fotografias ilustrando A) hábito arbustivo com crescimento indeterminado; B) aspecto das folhas de um indivíduo mostrando semelhança com as folhas de variedades cultivadas ilustradas nas Figuras 1 e 230
- Figura 4 - Mapa da distribuição do cultivo de variedades mansas e bravas de mandioca na região amazônica. O mapa foi compilado por Emperaire (2002) e modificado por Arroyo-Kalin (2010), e é baseado em dados etnográficos. O tamanho das barras é proporcional ao número de variedades cultivadas em cada região, sendo as variedades mansas representadas pela cor verde e as bravas pela cor laranja.....34
- Figura 5 - Esquema do método de *Restriction-site associated DNA sequencing* (RAD-seq). Amostras de DNA são submetidas à digestão com enzima de restrição e adaptadores P1 contendo diferentes sequências *barcodes* para cada indivíduo (retângulos vermelhos e azuis) são ligados aos fragmentos gerados. Uma etapa de fragmentação aleatória é seguida pela seleção de tamanho dos fragmentos. Adaptadores P2 são ligados aos fragmentos selecionados, e uma PCR seletiva para o enriquecimento das bibliotecas RAD é realizada. As sequências geradas para cada indivíduo podem ser alinhadas a um genoma de referência para o descobrimento de polimorfismos de nucleotídeos únicos (SNPs).....40
- Figura 6 - Mapa ilustrando as localidades (pontos) onde variedades de mandioca mansas e bravas foram coletadas junto às comunidades de agricultores tradicionais ao longo dos rios Madeira, Negro, Branco, Solimões e Amazonas. Também foram amostradas variedades do sul de Rondônia (rio Guaporé) e nordeste do Pará.....47
- Figura 7 - Esquema das etapas de amplificação e detecção de polimorfismo com marcadores microssatélites cloroplastidiais (cpSSR). A) amplificação de locos cpSSR por meio de PCR e detecção de polimorfismos em sequenciador Li-Cor 4300 *Analyzer*; B) os genótipos dos indivíduos foram determinados por meio da análise de géis virtuais, e a matriz de genótipos resultante foi utilizada para as análises estatísticas descritas no tópico 3.6.151
- Figura 8 - Esquema das etapas de amplificação e detecção de polimorfismo com marcadores microssatélites nucleares (ncSSR). A) amplificação de locos ncSSR por meio de PCR e detecção de polimorfismos em sequenciador ABI 3500*xl*; B) os genótipos dos indivíduos foram determinados por meio da análise de eletroferogramas, e a matriz de genótipos resultante foi utilizada para as análises estatísticas descritas no tópico 3.6.252

- Figura 9 - Mapa ilustrando as localidades (pontos) de onde as variedades de mandioca mansas e bravas incluídas nas análises de diversidade e estrutura genômica foram coletadas junto às comunidades de agricultores tradicionais ao longo dos rios Madeira, Negro, Branco, Solimões e Amazonas. Também foram amostrados indivíduos silvestres no estado de Rondônia 53
- Figura 10 - Esquema das etapas de descobrimento de polimorfismos de nucleotídeos únicos (SNPs) por meio da técnica de *Restriction-site associated DNA* (RAD-seq). A) e B) as bibliotecas RAD são preparadas e submetidas ao sequenciamento na plataforma *Illumina Genome Analyzer IIx*; C) As sequências obtidas foram separadas por indivíduos (“demultiplexagem”), e alinhadas a um genoma de referência; D) Marcadores SNPs foram descobertos nos locos RAD identificados pelo alinhamento, e a matriz de genótipos resultante foi utilizada para as análises estatísticas descritas no tópico 3.6.3..... 54
- Figura 11 - Rede de haplótipos observados em variedades cultivadas e silvestres de mandioca. Cada círculo representa um haplótipo, e o tamanho dos círculos é proporcional à frequência do haplótipo. Diferentes cores representam em A) mandiocas mansas, bravas e silvestres, e em B) mandiocas por rio ou região de coleta. Cada barra transversal representa uma mutação entre os haplótipos 64
- Figura 12 - Distribuição geográfica dos haplótipos observados em variedades cultivadas e silvestres de mandioca amostrados em diferentes municípios (pontos pretos) ao longo dos principais rios da bacia Amazônica brasileira, e no leste do estado do Pará. Cada haplótipo é identificado por uma cor diferente indicada na rede de haplótipos na caixa. Distribuição dos haplótipos considerando A) todos os indivíduos, B) mandiocas bravas e C) mandiocas mansas. O tamanho dos círculos não está relacionado ao número de amostras em cada local, mas alguns estão aumentados para melhor visualização..... 65
- Figura 13 - Estrutura genética sugerida pelo programa *Structure* para variedades de mandioca, com base em quatro locos cpSSR. A) Gráfico da estatística ΔK obtidos a partir de 10 simulações independentes para cada número K de agrupamentos. B) Gráficos de barras representando as probabilidades de ancestralidade de 596 indivíduos de mandioca considerando $K = 3$ e C) $K = 4$. Cada barra representa um indivíduo, e cores diferentes representam as probabilidades dos indivíduos de pertencerem a grupos genéticos distintos. As linhas pretas separam os indivíduos de acordo com os municípios. Asteriscos indicam os indivíduos silvestres 67
- Figura 14 - Análise Discriminante de Componentes Principais (DAPC) com base em quatro locos cpSSR para 596 indivíduos de mandioca. A) Gráfico de dispersão dos agrupamentos formados pelo método de *K-means*; B) Gráfico de densidade dos indivíduos de acordo com grupos de variedades cultivadas e silvestres ao longo da primeira função discriminante obtida na DAPC 68
- Figura 15 - Mapas mostrando o número total de alelos [coluna da esquerda, em células de 10 minutos (aproximadamente 18 Km) aplicando o método de *circular-neighborhood* de 0,45045 graus (aproximadamente 50 Km)], e a riqueza alélica (coluna da direita, em células de 0,45045 graus), com base na variação genética revelada por 14 marcadores ncSSR. Em A) todos os indivíduos, em B) apenas variedades de mandioca brava, e em C) apenas variedades de mandioca mansa 72

- Figura 16 - Gráficos de barra e interpolação espacial dos coeficientes de ancestralidade (q) para os agrupamentos identificados pela análise de estruturação genética obtida com o programa *Structure*. Mapas são mostrados para cada um dos K agrupamentos. A barra de escala à direita dos mapas representa a probabilidade dos indivíduos (pontos pretos) de pertencerem ao agrupamento. Estão representadas as análises com: A) todos os 596 indivíduos (asterisco indica os indivíduos silvestres), B) com 325 indivíduos de mandiocas bravas, e C) com 226 indivíduos de mandiocas mansas..74
- Figura 17 - Análise Discriminante de Componentes Principais (DAPC) com base em 14 locos ncSSR para 596 indivíduos de mandioca. A) Gráfico de dispersão dos sete agrupamentos formados pelo método de *K-means*; B) Gráfico de dispersão considerando três agrupamentos; C) Gráfico de densidade dos indivíduos de acordo com grupos de variedades cultivadas e silvestres ao longo da primeira função discriminante obtida na DAPC.....76
- Figura 18 - Estimativas de F_{ST} (diagonal inferior) e *heatmap* dos valores de F_{ST} (diagonal superior), com base em 14 locos ncSSR, entre pares de: A) variedades mansas, bravas e silvestres; B) entre grupos de variedades de acordo com os rios ou regiões; C) entre variedades mansas (M) e bravas (B) de acordo com rios ou regiões. Valores em negrito são significativos com intervalo de confiança de 95%. As cores variam do branco (menores valores de F_{ST}) ao azul escuro (maiores valores de F_{ST}).....77
- Figura 19 - Dendrograma mostrando os relacionamentos entre 596 indivíduos de variedades de mandiocas construído pelo método *Neighbor-Joining*, com base em distâncias de Cavalli-Sforza & Edwards (1967) a partir de 14 locos ncSSR78
- Figura 20 - Dendrogramas mostrando os relacionamentos entre grupos de mandiocas mansas e bravas de acordo com rios e regiões de coleta construído pelo método *Neighbor-Joining*, com base em distâncias de Cavalli-Sforza & Edwards (1967) a partir de 14 locos ncSSR. A) variedades agrupadas por rios e regiões de coleta. B) Variedades agrupadas por rios. Valores de *bootstrap* acima de 70% são apresentados79
- Figura 21 - Gráfico de dispersão ilustrando as médias das estimativas de heterozigosidades esperadas (H_E , azul) e observadas (H_O , preto) para os 5.871 marcadores SNPs detectados em 159 indivíduos de mandioca, sendo A) 19 indivíduos silvestres, B) 71 variedades bravas, e C) 69 variedades mansas80
- Figura 22 - Gráfico de dispersão ilustrando as médias das estimativas do coeficiente de endogamia (f) para os 5.871 marcadores SNPs detectados em 159 indivíduos de mandioca, sendo A) 19 indivíduos silvestres, B) 71 variedades bravas, e C) 69 variedades mansas81
- Figura 23 - Estimativas de F_{ST} (diagonal inferior) e *heatmap* dos valores de F_{ST} (diagonal superior), com base em 5.871 marcadores SNPs, entre pares de: A) variedades mansas, bravas e silvestres; B) entre grupos de variedades de acordo com os rios; C) entre variedades mansas (M) e bravas (B) de acordo com rios. Valores em negrito são significativos com intervalo de confiança de 95%. As cores variam do branco (menores valores de F_{ST}) ao azul escuro (maiores valores de F_{ST})84
- Figura 24 - Gráficos da distribuição das estimativas de F_{ST} e heterozigosidades para detecção de locos *outliers*, com base em 5.871 marcadores SNP, obtidos com os programas *Lositan* e *fdist2*. Cada ponto preto representa um loco SNP, e as linhas vermelha e

azul representam os limites superior e inferior de 99% do intervalo de confiança. Os locos que estão acima do limite superior são considerados hipoteticamente sob seleção diferencial. São considerados em A) variedades mansas, bravas e silvestres, e em B) variedades mansas e bravas 85

Figura 25 - Mapas mostrando o número total de alelos em células de 10 minutos (aproximadamente 18 Km) aplicando o método de *circular-neighborhood* de 0,45045 graus (aproximadamente 50 Km), com base na variação genômica revelada por 202 marcadores SNPs com comportamento *outlier*. Em A) 71 variedades de mandioca brava, e em B) 69 variedades de mandioca mansa 86

Figura 26 - Dendrogramas mostrando os relacionamentos entre 159 indivíduos de variedades de mandioca construído pelo método *Neighbor-Joining*, com base em distâncias de Cavalli-Sforza & Edwards (1967). São considerados em A) 5.213 marcadores SNPs com comportamento neutro, e B) 658 marcadores SNPs com comportamento *outlier* 87

Figura 27 - Dendrogramas mostrando relacionamentos entre mandiocas mansas e bravas de acordo com os rios, construído pelo método *Neighbor-Joining*, com base em distâncias de Cavalli-Sforza & Edwards (1967). São considerados em A) 5.213 marcadores SNPs com comportamento neutro, e B) 658 marcadores SNPs com comportamento *outlier*. Valores de *bootstrap* acima de 70% são apresentados..... 88

Figura 28 - Análises Discriminantes de Componentes Principais (DAPC) com marcadores SNPs para 159 indivíduos de mandioca cultivada e silvestre. Acima são apresentadas as análises com base em 5.213 marcadores SNPs com comportamento neutro, e abaixo são apresentadas as análises com base em 658 marcadores SNPs com comportamento *outlier*. Estão representados em A) e C) os agrupamentos formados pelo método de *K-means*, e em B) e D) os grupos de variedades mansas, bravas e silvestres..... 89

Figura 29 - Análises Discriminantes de Componentes Principais (DAPC) com marcadores SNPs para 140 indivíduos de variedades de mandioca mansas e bravas. Acima são apresentadas as análises com base em 5.669 marcadores SNPs com comportamento neutro, e abaixo são apresentadas as análises com base em 202 marcadores SNPs com comportamento *outlier*. Estão representados em A) e C) os agrupamentos formados pelo método de *K-means*, e em B) e D) os grupos de variedades mansas e bravas. 90

Figura 30 - Gráficos de barra representando as probabilidades de atribuição dos indivíduos aos agrupamentos considerados nas Análises Discriminante de Componentes Principais (DAPCs) com base em marcadores SNPs. DAPCs realizadas para variedades mansas, bravas e silvestres com base em: A) 5.213 marcadores SNPs com comportamento neutro, e B) com os 658 marcadores SNPs com comportamento *outlier*. DAPCs realizada para variedades mansas e bravas com base em: C) 5.669 marcadores SNPs com comportamento neutro, e D) com os 202 marcadores SNPs com comportamento *outlier* 91

Figura 31 - Análises Discriminantes de Componentes principais (DAPC) e interpolação espacial das probabilidades de atribuição aos agrupamentos identificados pelas DAPCs para 71 variedades de mandiocas bravas, com base em marcadores SNPs. Mapas são mostrados para cada um dos agrupamentos. A barra de escala à direita dos mapas representa a probabilidade dos indivíduos (pontos pretos) de pertencerem ao

agrupamento. Estão representadas as análises com: A) 5.669 marcadores SNPs com comportamento neutro, e B) 202 marcadores SNPs com comportamento *outlier*...92

Figura 32 - Análises Discriminantes de Componentes principais (DAPC) e interpolação espacial das probabilidades de atribuição aos agrupamentos identificados pelas DAPCs para 69 variedades de mandiocas mansas, com base em marcadores SNPs. Mapas são mostrados para cada um dos agrupamentos. A barra de escala à direita dos mapas representa a probabilidade dos indivíduos (pontos pretos) de pertencerem ao agrupamento. Estão representadas as análises com: A) 5.669 marcadores SNPs com comportamento neutro, e B) 202 marcadores SNPs com comportamento *outlier*...93

LISTA DE TABELAS

- Tabela 1 - Locais de coleta (Estados e Municípios) com os respectivos números de variedades de mandiocas mansas, bravas, não designadas (ND) e indivíduos silvestres amostrados. O georreferenciamento indica a sede dos municípios visitados.....48
- Tabela 2 - Estimativas de diversidade genética para os locos cpSSR polimórficos em 596 indivíduos de mandioca. A = número de alelos, I = Índice de Shannon, h = diversidade gênica61
- Tabela 3 - Estimativas de diversidade genética para os locos cpSSR, considerando os grupos de variedades cultivadas e os indivíduos silvestres, e os grupos de variedades cultivadas ao longo dos rios e regiões. $\bar{A}$ = número médio de alelos, I = Índice de Shannon, h = diversidade gênica61
- Tabela 4 - Haplótipos cloroplastidiais obtidos para variedades cultivadas e silvestres, com base em quatro locos cpSSR.....62
- Tabela 5 - Estimativas de diversidade haplotípica para variedades de mandioca cultivada e indivíduos silvestres, e grupos de variedades cultivadas ao longo dos rios. N = número de indivíduos, N_H = número de haplótipos, N_{eH} = número efetivo de haplótipos, I = Índice de Shannon, H_{Ehap} = diversidade haplotípica, π = diversidade nucleotídica..63
- Tabela 6 - Análise de variância molecular (AMOVA) baseada em quatro locos cpSSR para diferentes grupos de variedades de mandioca. G.L. = graus de liberdade. Valores em negrito indicam as estatísticas correspondentes a diferenciação entre grupos (F_{ST}) em cada AMOVA.....66
- Tabela 7 - Distribuição dos indivíduos de mandioca entre os agrupamentos formados pela Análise Discriminante de Componentes Principais (DAPC), baseado em quatro locos cpSSR, considerando os grupos de variedades, rios e regiões de coleta68
- Tabela 8 - Características e estimativas de diversidade genética para 14 locos ncSSR avaliados em 596 indivíduos de mandioca. Características: organização em sistemas *multiplex*, modificação fluorescente, e variação do tamanho de fragmentos amplificados em pares de base (pb). Estimativas de diversidade genética: número de alelos (A), heterozigosidades observada (H_O) e esperada (H_E), e estimativa do coeficiente de endogamia (f). *Significativo com intervalo de confiança de 95%.....69
- Tabela 9 - Estimativas de diversidade genética para 14 locos ncSSR avaliados em 596 indivíduos de mandioca, considerando variedades de mandioca cultivada e indivíduos silvestres, variedades cultivadas ao longo dos rios ou regiões, e grupos de variedades mansas e bravas ao longo dos rios ou regiões. Número de indivíduos (N), número de genótipos multilocos (MLGs), número médio de alelos ($\bar{A}$), riqueza alélica (A_R), número de alelos privados (A_p), heterozigosidades observada (H_O) e esperada (H_E), e estimativa do coeficiente de endogamia (f). *Significativo com intervalo de confiança de 95%70
- Tabela 10 - Distribuição dos indivíduos de mandioca entre os agrupamentos formados pela Análise Discriminante de Componentes Principais (DAPC), baseado em 14 locos ncSSR, considerando os grupos de variedades, rios e regiões de coleta75

- Tabela 11 - Análises de variância molecular (AMOVA) baseada na média de 14 locos ncSSR para diferentes grupos de variedades de mandioca. Valores em negrito indicam as estatísticas correspondentes a diferenciação entre grupos (F_{ST}) em cada AMOVA 76
- Tabela 12 - Estimativas de diversidade genômica para 5.871 marcadores SNPs avaliados em 159 indivíduos de mandioca, considerando variedades de mandioca cultivada e indivíduos silvestres, variedades cultivadas ao longo dos rios, e variedades mansas e bravas ao longo dos rios. Número de indivíduos (N), número total de alelos (A), riqueza alélica média por loco (A_R), heterozigosidades observada (H_O) e esperada (H_E), e coeficiente de endogamia (f). *Significativo com intervalo de confiança de 95% 82
- Tabela 13 - Análise de variância molecular (AMOVA) baseada na média de 5871 locos SNPs entre variedades mansas e bravas, entre indivíduos silvestres e variedades cultivadas de mandioca, e entre grupos de variedades de acordo com rios. Valores em negrito indicam as estatísticas correspondentes a diferenciação entre grupos (F_{ST}) em cada AMOVA. *Significativo com $p < 0,05$ 83
- Tabela 14 - Identificação, características genômicas e anotação funcional dos marcadores SNPs com comportamento neutro que apresentaram maior contribuição para a divergência de grupos de variedades de mandioca identificados pelas análises discriminantes de componentes principais (DAPCs) das Figuras 19 B e 20 B e que estavam associados à regiões transcritas do genoma parcial da mandioca (*Manihot esculenta* v4.1). São apresentados: os alelos de cada loco (SNP); a porcentagem da divergência entre grupos de variedades de mandioca explicada por cada loco nas DAPCs (%); a posição do loco SNP ao longo do genoma parcial da mandioca (*Scaffold* e Posição); o transcrito e a região gênica no qual o loco SNP está inserido; a anotação funcional baseada nas três classes de termos GO associada aos genes resultantes dos transcritos, depositados no banco de dados do *Joint Genome Institute* – JGI 96
- Tabela 15 - Identificação, características genômicas e anotação funcional dos marcadores SNPs com comportamento *outlier* que apresentaram maior contribuição para a divergência de grupos de variedades de mandioca identificados pelas análises discriminantes de componentes principais (DAPCs) das Figuras 19 D e 20 D e que estavam associados à regiões transcritas do genoma parcial da mandioca (*Manihot esculenta* v4.1). São apresentados: os alelos de cada loco (SNP); a porcentagem da divergência entre grupos de variedades de mandioca explicada por cada loco nas DAPCs (%); a posição do loco SNP ao longo do genoma parcial da mandioca (*Scaffold* e Posição); o transcrito e a região gênica no qual o loco SNP está inserido; a anotação funcional baseada nas três classes de termos GO associada aos genes resultantes dos transcritos, depositados no banco de dados do *Joint Genome Institute* – JGI 98

1 INTRODUÇÃO

Dentre os cultivos que tiveram origem na Amazônia, a mandioca (*Manihot esculenta* ssp. *esculenta* Crantz) é o cultivo alimentício domesticado que alcançou a maior importância no mundo. Atualmente, é o oitavo cultivo mais importante, atrás de cana-de-açúcar, milho, arroz, trigo, batata, beterraba e soja (FAOStat, 2013). É um arbusto perene, considerado bem adaptado a solos ácidos e de baixa fertilidade, sendo cultivado ao redor de toda a região tropical em agroecossistemas com estas características. A mandioca produz raízes que acumulam grandes quantidades de amido, sendo a principal fonte de carboidratos para mais de 800 milhões de pessoas no mundo, principalmente nos países tropicais subdesenvolvidos (LEBOT, 2009). Análises moleculares e evidências morfológicas apontam que a domesticação da mandioca ocorreu a partir de populações de *Manihot esculenta* ssp. *fabellifolia*, entre os estados de Mato Grosso, Rondônia, Acre, e áreas adjacentes na Bolívia, no sudoeste da bacia Amazônica (ALLEM, 2002; OLSEN; SCHAAL, 1999). Após sua domesticação inicial, pressões seletivas divergentes deram origem aos dois grandes grupos de variedades atualmente cultivadas: a mandioca mansa e a mandioca brava (ELIAS et al., 2004; MÜHLEN; MARTINS; ANDO, 2000). As variedades mansas apresentam menores quantidades de substâncias tóxicas, sendo comumente cultivadas nas bordas da bacia amazônica (nas cabeceiras dos principais rios) e ao longo dos principais rios da Amazônia ocidental, incluindo os rios Ucayali e Marañón no Peru, na região periférica do sul da Amazônia, e também na Mesoamérica. As variedades mansas são cultivadas também em menor escala onde as variedades bravas são o cultivo principal. Por sua vez as variedades bravas apresentam alto conteúdo de substâncias tóxicas e seu cultivo está associado às margens dos principais rios da bacia amazônica e ao litoral oceânico entre as Guianas e o Maranhão (EMPERAIRE, 2001). Embora existam áreas de sobreposição na distribuição de variedades mansas e bravas, as distribuições dos grupos de variedades podem ser resultantes de processos de dispersão independentes, e processos limitados de intercâmbio de variedades entre populações humanas ao longo da história da domesticação deste cultivo (EMPERAIRE, 2001). Grande parte do cultivo de mandioca, especialmente na Amazônia, é realizada em sistemas de agricultura tradicional (EMPERAIRE, 2005). Estes sistemas estão associados ao manejo e a manutenção de uma grande variabilidade genética existente nas variedades presentes nas roças de mandioca (MARTINS, 2001).

A compreensão da variabilidade e estrutura genética das populações cultivadas de uma espécie é fundamental para orientar a prospecção de seus recursos genéticos, bem como para entender a história evolutiva dessas populações ao longo do processo de domesticação pré e

pós-conquista europeia (CLEMENT et al., 2010; NASS, 2001; VENCovsky et al., 2007). Esta informação está começando a ser gerada na Amazônia para diversas espécies importantes, como é o caso da mandioca. Muita informação pôde ser obtida rapidamente com o uso de marcadores moleculares de várias classes. De fato, diversos estudos envolvendo análises genéticas de variedades de mandioca cultivada tradicionalmente vêm sendo realizados e permitiram, por exemplo, a elucidação do centro de origem do cultivo (OLSEN; SCHAAL, 1999), a detecção de zonas de hibridação com parentes silvestres (DUPUTÍÉ et al., 2007), o aumento da compreensão da variabilidade e estrutura genética de populações domesticadas (PERONI; KAGEYAMA; BEGOSSI, 2007), bem como aumento do entendimento da dinâmica evolutiva da mandioca sob domesticação (ALVES-PEREIRA et al., 2011). Entretanto, poucos destes estudos foram voltados para a análise de como a variabilidade genética dentro do cultivo está organizada ao longo de sua distribuição geográfica, embora alguns tentaram abordar a questão mesmo com amostragens disjuntas (ELIAS et al., 2004; MÜHLEN; MARTINS; ANDO, 2000).

Desta forma, entender como a variabilidade genética de conjuntos de variedades de mandioca cultivadas tradicionalmente ao longo dos principais eixos fluviais da bacia Amazônica está organizada foi o objetivo geral deste estudo. Para tanto foram empregados marcadores microssatélites nucleares e cloroplastidiais, além da detecção de SNPs através da metodologia de RAD-seq (*Restriction-site Associated DNA Sequencing*). Foram avaliadas: (i) a estrutura genética contemporânea das variedades tradicionais de mandioca; (ii) a estrutura genética histórica com a reconstrução das relações matrilineares entre os grupos de variedades mansas e bravas; e (iii) informações inéditas sobre as bases genômicas para a diferenciação entre os grupos de variedades de mandioca mansas e bravas. Estas informações foram interpretadas em termos da dispersão e difusão do cultivo a partir do seu centro de origem, na porção sudoeste da bacia Amazônica, para outras localidades da Amazônia brasileira através dos principais rios amazônicos. O entendimento da história de dispersão da mandioca tem importância prática, uma vez que pode permitir a identificação de áreas com concentração de variabilidade genética que poderão orientar futuras prospecções e esforços de conservação dos recursos genéticos da mandioca, tanto *in situ* como *ex situ*, e que também podem ser utilizados para o melhoramento do cultivo.

1.1 Objetivos

1.1.1 Objetivo geral

Avaliar como a variabilidade genética e genômica presente em variedades mansas e bravas de mandioca (*Manihot esculenta* Crantz), cultivadas tradicionalmente por comunidades de agricultores familiares, está organizada ao longo dos principais eixos fluviais da bacia amazônica no Brasil, utilizando marcadores moleculares e dados genômicos para identificar possíveis padrões de dispersão do cultivo na região.

1.1.2 Objetivos específicos

1. Analisar os padrões históricos de distribuição da variabilidade haplótipica entre linhagens matrilineares de variedades de mandioca cultivadas tradicionalmente ao longo dos principais eixos fluviais da bacia amazônica com marcadores microssatélites cloroplastidiais (cpSSR);
2. Analisar a estrutura e diversidade genética contemporânea de variedades de mandioca cultivadas tradicionalmente ao longo dos principais eixos fluviais da bacia amazônica com marcadores microssatélites nucleares (ncSSR);
3. Analisar a variação genômica de variedades de mandioca cultivadas tradicionalmente ao longo dos principais eixos fluviais da bacia amazônica com marcadores SNPs detectados pela técnica RAD-seq (*Restriction-site Associated DNA sequencing*);
4. Avaliar com a abordagem de genômica de populações os perfis genômicos das variedades mansas e bravas, discriminando marcadores associados à diferenciação entre os grupos de variedades, bem como aqueles associados a possíveis padrões de dispersão distintos;
5. Gerar informações sobre prováveis rotas de dispersão do cultivo a partir do seu centro de origem no sudoeste para outras regiões da bacia amazônica brasileira, com base nos resultados de estruturação genética contemporânea, histórica e perfis genômicos.

1.2 Hipóteses

1. Os marcadores cpSSR são eficientes para a detecção de polimorfismos em mandioca e a reconstrução do relacionamento das linhagens matrilineares de variedades de mandioca e revelarão padrões de relacionamentos distintos para os grupos de mandioca mansas e bravas;
2. Existe estruturação genética e genômica entre variedades de mandioca mansas e bravas

(como descrito na literatura, em outras regiões), bem como entre diferentes localidades ao longo dos principais rios da bacia amazônica no Brasil;

3. Existem locos SNPs com comportamento *outlier* que possivelmente estejam associados à diferenciação entre variedades mansas e bravas e que podem ser detectados pela técnica de RAD-seq;
4. As variedades de mandioca mansas e bravas apresentam padrões distintos de organização da diversidade genética que podem ser explicados por eventos diferenciados de dispersão das variedades durante a história de domesticação do cultivo.

2 REVISÃO BIBLIOGRÁFICA

2.1 Aspectos gerais da mandioca

A mandioca (*Manihot esculenta* Crantz, Euphorbiaceae) é a única espécie alimentícia cultivada do gênero (ROGERS; APAN, 1973). É uma planta arbustiva alcançando entre 1 e 4 metros de altura (ALVES, 2002) (Figura 1). A mandioca apresenta flores masculinas e femininas agrupadas em inflorescências terminais do tipo panícula. Insetos de pequeno porte são responsáveis pela polinização cruzada, embora a autopolinização também possa ocorrer. O fruto é uma cápsula trilocular e possui uma semente com carúncula (pequena excrescência carnosa no tegumento) em cada lóculo (ALVES, 2002; LEBOT, 2009). As plantas propagadas vegetativamente (por estacas) apresentam várias raízes adventícias. Embora sejam os principais órgãos de reserva, apenas algumas raízes acumulam amido e se tornam tuberosas. Por outro lado, plantas oriundas de sementes desenvolvem apenas uma raiz pivotante que acumula amido. Dependendo do número de nós, as estacas dão origem de um a três caules lenhosos que apresentam nós protuberantes e padrão de ramificação variável. As folhas são alternadas, geralmente glabras, lobuladas (em geral de três a nove, ocasionalmente até 11 lóbulos) (ALVES, 2002; LEBOT, 2009).

Figura 1 - Aspectos morfológicos da planta de mandioca (*Manihot esculenta* ssp. *esculenta*). Fotografias ilustrando A) hábito arbustivo de uma variedade de mandioca brava; B) aspecto das raízes logo após a colheita; C) raízes descascadas antes do processamento de detoxificação (Fotografias por André Junqueira)

Estima-se que hajam aproximadamente 20.000 acessos de mandioca representados em bancos de germoplasma e nas coleções nucleares do CIAT (*Centro Internacional de Agricultura Tropical*), IITA (*International Institute of Tropical Agriculture*) e em mais de 45 países no mundo todo (BONIERBALE et al., 1997). A mandioca cultivada exibe uma grande variedade de características morfológicas (Figura 2). Esta variação pode ser resultante, por exemplo, da ocorrência de hibridização intraespecífica (entre variedades diferentes), e interações genótipo-ambiente (GxA) (ALVES, 2002). Na maioria dos países em que a mandioca é cultivada os agricultores selecionaram variedades bem adaptadas às condições locais. Frequentemente observa-se que variedades selecionadas em outras localidades apresentam performance inferior às localmente selecionadas (LEBOT, 2009). Outros estudos demonstraram que a produtividade de variedades de mandioca pode ser fortemente afetada por interações GxA (CALLE et al., 2005, JAMARILLO et al., 2005). Interações GxA também já foram documentadas para a composição de nutrientes e potencial cianogênico de variedades (BOKANGA et al., 1994; BURNS et al. 2012). Adicionalmente, pelo menos parte da variação morfológica pode ser explicada também por hibridização entre variedades cultivadas e outras espécies de *Manihot* (LEBOT, 2009). A possibilidade de hibridização interespecífica pode ser evidenciada pela proposição de um *pool* gênico primário constituído por três espécies silvestres, e um *pool* gênico secundário constituído por outras 13 espécies de *Manihot* para a mandioca (ALLEM et al., 2001). Isto também é refletido na ausência de um grupo monofilético de variedades cultivadas em uma análise filogenética realizada para algumas espécies do gênero (CHACÓN et al., 2008, DUPUTIE et al., 2011).

Figura 2 - Variação morfológica presente em variedades cultivadas representada por: variedades de mandioca brava cultivada em A) solos antropogênicos; B) Latossolos; e variedade mansa cultivada em C) solo de várzea. São perceptíveis variações na cor e formato das folhas e na cor dos pecíolos. (Fotografias por André Junqueira)

Os principais centros de pesquisa de mandioca CIAT e IITA, e outros institutos de pesquisas no Brasil, Cuba, Índia, Tailândia, Filipinas, entre outros países, conduzem programas de melhoramento de mandioca (LEBOT, 2009). Entretanto, os objetivos dos programas de melhoramento variam entre os países, e de acordo com o propósito, para uso industrial, ou consumo e subsistência humana (LEBOT, 2009). Por exemplo, para uso industrial procura-se o melhoramento da produtividade (de amido e matéria seca) por unidades de área e tempo. Por outro lado, para a subsistência os objetivos do melhoramento são mais complexos e envolvem preferências por variedades mansas ou bravas e estabilidade. Embora existam cultivares modernos resultantes do melhoramento do cultivo e haja alguma demanda de mercado para fabricação de alguns produtos, grande parte do cultivo se dá por meio da agricultura tradicional com pouco ou nenhum emprego de mecanização e insumos; o que é especialmente verdadeiro na Amazônia. Desta forma, a mandioca possui grande importância para a segurança alimentar de uma grande parcela de agricultores familiares em países subdesenvolvidos. De fato, ela é a principal fonte de carboidratos para cerca de 800 milhões de pessoas (LEBOT, 2009).

2.2 Domesticação do cultivo

A origem da mandioca como cultivo sempre foi muito debatida. O gênero *Manihot* possui 98 espécies distribuídas por toda a região Neotropical desde o México até o norte da Argentina. Estimativas baseadas em análises filogenéticas sugerem que o gênero surgiu a cerca de 6,6 milhões de anos (CHACÓN et al., 2008), e provavelmente tem origem Mesoamericana, onde se diversificou antes de colonizar a América do Sul (DUPUTÍÉ et al., 2011). Dentro desta distribuição são reconhecidos dois centros de diversidade específica do gênero, o primeiro no Brasil, com cerca de 80 espécies, e o segundo no México, com cerca de 17 espécies (SCHAAL; OLSEN; CARVALHO, 2006). A grande sobreposição de caracteres morfológicos entre as espécies ao longo da distribuição do gênero torna difícil a discriminação confiável de espécies baseada apenas na variação fenotípica. Esta sobreposição de caracteres morfológicos sugerem uma recente e explosiva diversificação do gênero, e os baixos níveis de variação observada em sequências de DNA parecem corroborar esta hipótese (CHACÓN et al., 2008). Isto contribuiu para que a mandioca fosse considerada por muito tempo um cultígeno (uma planta domesticada sem parente silvestre) e que teria se originado provavelmente de vários eventos de introgressão entre diferentes espécies silvestres (SCHAAL; OLSEN; CARVALHO, 2006).

Baseado em similaridades fenotípicas especulou-se que as espécies *Manihot aesculifolia* e *M. carthaginensis*, ambas de ocorrência na América Central, poderiam ser parentes silvestres

da mandioca (ROGERS; APAN, 1973). Allem (1994, 2002), também baseado em caracteres fenotípicos, propôs que a mandioca era composta por três subespécies: *Manihot esculenta* ssp. *esculenta* (a forma cultivada), ssp. *flabellifolia* (o parente silvestre mais próximo, Figura 3) e ssp. *peruviana*. Estas três subespécies juntamente com *Manihot pruinosa* formariam o *pool* gênico primário da mandioca. Allem (1994) propôs ainda que a domesticação ocorrera na região do Planalto Central, a principal área do centro de diversidade brasileiro, onde ocorrem 53 espécies de *Manihot*.

Figura 3 - *Manihot esculenta* ssp. *flabellifolia*. Fotografias ilustrando A) hábito arbustivo com crescimento indeterminado; B) aspecto das folhas de um indivíduo mostrando semelhança com as folhas de variedades cultivadas ilustradas nas Figuras 1 e 2

A origem do cultivo começou a ser elucidada no final da década de 90, quando evidências morfológicas e moleculares sugeriram que a mandioca foi domesticada em um único evento a partir de populações de *M. esculenta* ssp. *flabellifolia* na região sudoeste da bacia amazônica. Tais evidências provieram de estudos da diversidade genética de variedades cultivadas de *M. esculenta* ssp. *esculenta* e de populações de *M. esculenta* ssp. *flabellifolia* e *M. pruinosa* (potencialmente hibridizante) utilizando três tipos diferentes de marcadores moleculares [sequenciamento do gene nuclear de cópia única gliceraldeído 3-fosfato desidrogenase (*G3pdh*), polimorfismos de nucleotídeos únicos (SNPs) e microssatélites (SSRs)] (OLSEN, 2004; OLSEN; SCHAAL, 1999, 2001). Estes estudos demonstraram que *M. pruinosa* não contribuiu para o *pool* gênico da mandioca cultivada e que a variabilidade genética encontrada nas variedades cultivadas era um subconjunto da variabilidade genética presente em populações de *M. esculenta* ssp. *flabellifolia* que ocorrem nos estados brasileiros de Mato Grosso, Rondônia, Acre e áreas adjacentes na Bolívia (OLSEN, 2004). Recentemente, Léotard et al. (2009) avaliaram a variação no gene *G3pdh* com uma amostragem mais abrangente da

atual distribuição de *M. esculenta* ssp. *flabellifolia*, e ainda outras espécies de *Manihot* potencialmente hibridizantes com a mandioca. Estes autores validaram os resultados encontrados anteriormente.

Apesar disso, muita discussão ainda existe acerca da origem botânica da mandioca. Alguns autores questionam a evidência genética no sentido de que estes estudos foram temporalmente e espacialmente limitados, tornando difícil o entendimento da domesticação do cultivo, pelo menos sob ponto de vista paleo-arqueológico (PERRY, 2002). Mesmo estudos filogenéticos levantam esta discussão ao ressaltarem que os outros estudos genético-moleculares não envolveram uma gama de espécies suficientemente ampla para obter uma conclusão definitiva (DUPUTIÉ et al., 2011). Piperno e Holst (1998) verificaram que existem diferenças nos padrões de formação e morfologia de grânulos de amido entre mandiocas provenientes do Peru e do Panamá. Estes tipos distintos podem ser representativos de diferentes táxons, sendo que as amostras do Panamá parecem ser mais relacionadas à *M. esculenta* ssp. *flabellifolia*. Dentre outros, tais argumentos suportariam a hipótese de eventos múltiplos de domesticação da mandioca (UGENT et al., 1986).

2.3 Divergência entre grupos de variedades cultivadas

O processo de domesticação da mandioca causou mudanças significativas no hábito das variedades cultivadas em relação à *M. esculenta* ssp. *flabellifolia*. O favorecimento de características que facilitam a reprodução clonal é um exemplo marcante da síndrome de domesticação da mandioca (McKEY et al., 2012). Nota-se, por exemplo, uma redução no número de ramificações do caule da planta, além do aumento do diâmetro do caule, o que resulta em estacas usadas na reprodução clonal com maior reserva de nutrientes e menos sujeitas ao ressecamento (JENNINGS, 1995; SCHAAL; OLSEN; CARVALHO, 2006). Uma curiosa síndrome de domesticação se refere à morfologia das plântulas de variedades cultivadas e o parente silvestre da mandioca. Enquanto plântulas de *M. esculenta* ssp. *flabellifolia* apresentam germinação hipógea, as plântulas de mandioca cultivada têm germinação epígea (McKEY et al., 2012; PUJOL et al., 2005). A permanência dos cotilédones abaixo do nível do solo garante reservas para o crescimento de indivíduos silvestres em caso de herbivoria ou distúrbios ambientais. De outra forma, em detrimento da capacidade de regeneração em caso de danos à parte aérea da plântula, a exposição dos cotilédones acima do nível do solo faz com que estes tornem-se estruturas fotossintetizantes, colaborando para o desenvolvimento e crescimento mais rápido das variedades cultivadas (McKEY et al., 2012; PUJOL et al., 2005).

Outra característica marcante do processo de domesticação foi o desenvolvimento de raízes capazes de estocar grandes quantidades de amido. Após a sua domesticação inicial, pressões seletivas divergentes deram origem a dois grandes grupos de variedades atualmente cultivadas: a mandioca mansa (aipim, mandioca de mesa, macaxeira) e a mandioca brava (mandioca, mandioca amarga) (ELIAS et al., 2004; MÜHLEN; MARTINS; ANDO, 2000). Esta separação é baseada no potencial cianogênico das raízes que acumulam linamarina e lotaustralina, substâncias que são hidrolisadas a ácido cianídrico quando os tecidos da planta são danificados e que tornam as raízes de mandioca muito tóxicas. As variedades mansas apresentam menores quantidades de substâncias cianogênicas (< 100 mg/Kg de polpa da raiz) e podem ser consumidas com processamento básico (descascamento e cozimento, ou mesmo cruas). As variedades bravas possuem maiores quantidades de substâncias cianogênicas (> 100 mg/Kg de polpa da raiz) e exigem um processamento prévio de detoxificação antes de seu consumo na forma de farinha, beiju, tucupi, bebidas fermentadas, entre outros. Embora esta separação seja dicotômica, observa-se que a variação no conteúdo de substâncias tóxicas é contínua entre as variedades de mandioca e de seus parentes silvestres. Por conseguinte, a hipótese de que a mandioca brava e a mandioca mansa são resultado de seleções independentes a partir de um ancestral com nível médio de toxidez parece ser a mais plausível (McKEY; BECKERMAN, 1993).

Recentemente Arroyo-Kalin (2010) propôs a hipótese de que a distinção de variedades bravas e mansas estaria relacionada à intensificação da agricultura praticada pelas sociedades sedentárias na Amazônia do período pré-Colombiano. Variedades mansas teriam surgido primeiro como resultado da seleção para menor toxidez realizada por horticultores incipientes em ambientes próximos às unidades familiares. Devido à natureza nômade das primeiras sociedades, as variedades mansas teriam sido dispersas amplamente na Amazônia. Com o início do período Formativo, em que houve o aumento do sedentarismo nas Américas, houve um aumento da necessidade por alimentos, que começaram a ser cultivados em ambientes mais afastados das unidades familiares (roças). Tais ambientes são mais suscetíveis a pragas e doenças, o que favoreceria a seleção de variedades mais tóxicas por seleção natural e que tivessem maior produtividade por seleção humana. Desta forma, as variedades de mandioca bravas surgiram após a intensificação da agricultura e do concomitante desenvolvimento de tecnologias que pudessem ser empregadas em sua detoxificação, garantindo a segurança alimentar das primeiras sociedades sedentárias amazônicas. Em um estudo recente, Peña-Venegas et al. (2014) relatam o mito de etnias da Amazônia colombiana para a origem das variedades cultivadas de mandioca, no qual as primeiras roças de mandioca teriam sido

estabelecidas em solos antropogênicos (*Amazonian Dark Earths*, ADE). Solos antropogênicos são caracterizados por possuírem fertilidade bem mais elevada do que os Latossolos e Argissolos amazônicos, e são considerados legado do período de intensificação da agricultura por civilizações amazônicas pré-Colombianas (ARROYO-KALIN, 2009; WOODS; DENEVAN, 2009). Desta forma, o mito relatado por Peña-Venegas et al. (2014) pode ser entendido como evidência da hipótese levantada por Arroyo-Kalin (2010).

Em mandioca parece existir uma forte seleção para a obtenção de variedades com alta toxidez, já que as variedades bravas são cultivadas em maior escala em algumas regiões, como na Amazônia (CLEMENT et al., 2010). Na região amazônica se observa que o cultivo de variedades bravas está associado às margens dos principais rios da bacia amazônica e ao litoral Atlântico da América do Sul, principalmente entre a Guiana e o Maranhão. Por sua vez o cultivo de variedades mansas é mais disperso, sendo comumente cultivada em menor escala onde as variedades bravas e outras espécies são o cultivo principal. Existe a tendência de que a mandioca mansa seja o cultivo principal nas cabeceiras dos principais rios da bacia amazônica e também na porção mais ocidental da Amazônia, inclusive ao longo de rios importantes como o Ucayali e Marañon no Peru, e também na Mesoamérica (CLEMENT et al., 2010; McKEY; BECKERMAN, 1993). Apesar de existir sobreposição na distribuição das mandiocas mansas e bravas suas áreas de cultivo são um tanto diferentes (Figura 4). Segundo Emperaire (2002) este tipo de distribuição e indícios de natureza linguística e genética sugerem a provável existência de mais de um foco de diversificação da mandioca, ou seja, é provável que as variedades mansas e bravas possuam centros de diversificação diferentes. Emperaire (2002) especula ainda que estes padrões de dispersão diferentes resultaram de processos limitados de contato e intercâmbio entre os dois grupos de variedades ao longo da história do processo de domesticação do cultivo. Desta forma, o estudo da organização da variabilidade genética, tanto a de cunho recente, como a de cunho histórico revelado por linhagens haplotípicas matrilineares, e variabilidade genômica de variedades de mandioca poderia ser essencial para o entendimento dos padrões de difusão e dispersão ao longo da região amazônica dos grupos de variedades do cultivo após sua domesticação.

Figura 4 - Mapa da distribuição do cultivo de variedades mansas e bravas de mandioca na região amazônica. O mapa foi compilado por Emperaire (2002) e modificado por Arroyo-Kalin (2010), e é baseado em dados etnográficos. O tamanho das barras é proporcional ao número de variedades cultivadas em cada região, sendo as variedades mansas representadas pela cor verde e as bravas pela cor laranja

2.4 Dinâmica evolutiva de mandioca sob domesticação

Embora a propagação vegetativa seja um resultado marcante do processo de domesticação, a mandioca não perdeu sua capacidade de reprodução sexual. A fertilidade é variável entre as variedades, e a produção de flores e sementes se dá de maneira irregular (LEBOT, 2009). A reprodução sexual juntamente com a principal forma de cultivo (*swidden agriculture*, agricultura itinerante) exerce um papel fundamental na dinâmica evolutiva da mandioca como cultivo (EMPERAIRE, 2005). Nos sistemas de agricultura itinerante as roças são estabelecidas pela derrubada e queima de áreas florestais de vegetação secundária, menos frequentemente vegetação primária, e são utilizadas por períodos que variam de 1 a 4 anos. Após seu uso estas roças são deixadas em pousio durante um período que pode durar até 20 anos ou mais, enquanto uma nova vegetação secundária cresce e a fertilidade do solo é restabelecida. Após o período de pousio a roça é restabelecida a partir do corte e queima da nova vegetação secundária (ELIAS; RIVAL; McKEY, 2000; MARTINS, 2001; McKEY; BECKERMAN, 1993).

As variedades cultivadas nas roças podem ser colhidas conforme a necessidade, sendo possível que algumas plantas desenvolvam flores, que são polinizadas por insetos pequenos (LEBOT, 2009). Uma vez produzidas, as sementes são dispersas por explosão dos frutos

deiscentes (autocoria) e por formigas (mirmecoria) que se alimentam da carúncula presente nas sementes (ELIAS; McKEY, 2000). As formigas enterram as sementes, que passam a ser constituintes do banco de sementes do solo e que permanecem viáveis, mesmo quando as roças são deixadas em pousio por períodos longos. A dormência das sementes é quebrada pela elevação da temperatura causada pela queima da vegetação antes do estabelecimento de um novo ciclo de manejo de roças (PERONI; HANAZAKI, 2002; PUJOL et al., 2002). Desta forma, a germinação das sementes e o crescimento de plântulas produzidas sexualmente (voluntárias) ocorrem ao mesmo tempo em que uma nova roça está sendo estabelecida. Assim, existe a possibilidade de que plântulas voluntárias cresçam junto com plantas propagadas vegetativamente (DUPUTIÉ et al., 2009; MARTINS, 2001; PUJOL et al., 2007). Em geral, os agricultores conseguem distinguir as plântulas voluntárias daquelas propagadas vegetativamente. Consciente ou inconscientemente (nos casos em que os agricultores não reconhecem plântulas voluntárias) os agricultores podem permitir que tais plântulas se desenvolvam. Na época da colheita os agricultores podem avaliar se estas plantas oriundas de reprodução sexual serão selecionadas para propagação vegetativa (ELIAS; RIVAL; McKEY, 2000; McKEY; BECKERMAN, 1993; RIVAL; McKEY, 2008). Caso plantas oriundas de sementes sejam utilizadas para a propagação vegetativa elas podem ser incluídas em variedades já existentes e presentes nas roças, ou podem passar a formar uma nova variedade (BOSTER, 1985, 1986; PERONI, 1998). Portanto, a incorporação de plântulas oriundas de sementes colabora para amplificar a diversidade genética intraespecífica do cultivo (DUPUTIÉ et al., 2009; ELIAS et al., 2001; MARTINS, 2001; McKEY; BECKERMAN, 1993; SAMBATTI; MARTINS; ANDO, 2001). Ao amplificar a diversidade genética a incorporação de plântulas voluntárias ajuda a manter o vigor vegetativo das variedades locais, já que existe correlação entre o tamanho, heterozigosidade e sobrevivência das plantas (PUJOL; McKEY, 2006). Além disso, o processo de incorporação de plântulas voluntárias pode colaborar para a diminuição da susceptibilidade à pragas e doenças (ELIAS et al., 2004; PUJOL; DAVID; McKEY, 2005).

A importância da prática de incorporação de plântulas voluntárias ao material genético foi confirmada por meio de análises genético-moleculares que detectaram alta diversidade genética em variedades locais (DUPUTIÉ et al., 2009; ELIAS et al., 2001; PERONI; KAGEYAMA; BEGOSSI, 2007). Além disso a incorporação de plântulas voluntárias resulta em variedades locais polifiléticas, nas quais existe a preponderância de um clone e mais um conjunto de clones similares em termos morfológicos, mas diferentes em termos genéticos (PERONI, 1998, 2004; PUJOL et al., 2007). O estudo da 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 da base genética local (EMPERAIRE; PERONI, 2007; PERONI; HANAZAKI, 2002).

2.5 Marcadores moleculares microsatélites

Marcadores moleculares são quaisquer fenótipos moleculares provenientes de quaisquer segmentos expressos ou não de DNA que são herdados geneticamente e se prestam à diferenciação de um ou mais indivíduos (FERREIRA; GRATTAPAGLIA, 1998; MILACH, 1998). Entre as vantagens da utilização de marcadores moleculares em relação aos marcadores morfológicos estão o fato de serem mais numerosos e detectarem maior quantidade de polimorfismos. Além disso, comportam-se como caracteres de herança simples e previsível, não sendo afetados por variações no meio ambiente. Desta forma, os marcadores moleculares permitem quantificar a diversidade genética, seguir o movimento de indivíduos, estimar endogamia, caracterizar novas espécies e avaliar os padrões históricos de dispersão da espécie ou determinadas populações em questão (FREELAND, 2005).

Existem várias técnicas disponíveis para a obtenção de marcadores moleculares, cada uma utilizando uma estratégia particular para detectar polimorfismos. Os tipos de marcadores moleculares diferem entre si também com relação a características tais como a abundância no genoma, os níveis de polimorfismo detectados, especificidade de locos, reprodutibilidade, requerimentos técnicos e custo (MONDINI; NOORAMI; PAGNOTTA, 2009).

Marcadores microsatélites nucleares (ncSSR) são trechos de DNA consistindo em unidades de um a seis pares de bases repetidos em tandem e que podem ser encontrados ao longo de todo o genoma da maioria dos eucariotos (POWELL; MACHRAY; PROVAN, 1996). A detecção do polimorfismo é realizada pela amplificação dos locos microsatélites por meio de PCR a partir do DNA genômico total, usando um par de iniciadores que flanqueiam um loco microsatélite específico.

Os ncSSR são altamente polimórficos, possuem natureza multialélica, são co-dominantes, são de fácil detecção (por meio de PCR) e análise, são abundantes no genoma e sua detecção requer pequenas quantidades de DNA. Diferentes melhorias nos métodos de descobrimento e análises de ncSSR tem reduzido os custos e impulsionado a utilização destes marcadores (GUICHOUX et al., 2011). Além disso, possuem grande reprodutibilidade e as sequências iniciadoras podem ser compartilhadas entre diferentes grupos de pesquisa (GOLDSTEIN; POLLOCK, 1997; POWELL; MACHRAY; PROVAN, 1996). Estas propriedades tornaram os microsatélites um dos marcadores mais populares em estudos de

genética de populações, mapeamento, testes de paternidade, entre outros (SCHLÖTTERER, 2004).

Além de ocorrerem no genoma nuclear, os microssatélites também são encontrados no genoma do cloroplasto (cpSSR) e, para algumas espécies, também no genoma mitocondrial (AVISE, 2004; FREELAND, 2005). Os cpSSR, inicialmente descritos por Powell et al. (1995a, 1995b), são tipicamente compostos por sequências curtas de mononucleotídeos, ocorrendo em geral em regiões não codificantes, e apresentando variação intraespecífica no número de repetições (EBERT; PEAKALL, 2009). Embora sejam superficialmente semelhantes aos ncSSR, os cpSSR são caracterizados pela haploidia e, com algumas exceções, pela falta de recombinação. Ao contrário dos marcadores nucleares, os marcadores cloroplastidiais são necessariamente ligados, e consequentemente um conjunto de vários cpSSR pode ser tratado como um haplótipo que possui sua própria história evolutiva (SCHAAL et al., 1998). Estas características aliadas à herança uniparental tendem a reduzir o tamanho populacional efetivo do genoma do cloroplasto em comparação com o genoma nuclear. Isto faz com que a variação do genoma cloroplastidial seja comparativamente mais sensível a efeitos genéticos, como a deriva. Desta forma, espera-se que os marcadores cpSSR sejam mais efetivos na detecção da estruturação e subdivisão populacional do que os ncSSR (PETIT et al., 2005; SCHAAL et al., 1998). A variação revelada por cpSSR é também especificamente importante para estudos de filogeografia, além de complementar as informações que podem ser obtidas com marcadores ncSSR.

Na literatura atualmente existem 186 locos ncSSR descritos para mandioca (CHAVARRIAGA-AGUIRRE et al., 1998; MBA et al., 2001). Apesar do genoma completo do cloroplasto de mandioca ter sido descrito recentemente (DANIELL et al., 2008), ainda não foram desenvolvidos marcadores cpSSR para a espécie. Os marcadores ncSSR vêm sendo utilizados com diversos objetivos, como para quantificar a diversidade genética presente nas amostras de bancos de germoplasma e coleções nucleares (CHAVARRIAGA-AGUIRRE et al., 1998, 1999; ELIAS et al., 2004; FREGENE et al., 2003; SIQUEIRA et al., 2009). Marcadores ncSSR foram úteis no auxílio à determinação do centro de origem do cultivo (OLSEN, 2004; OLSEN; SCHAAL, 2001). Peroni, Kageyama e Begossi (2007) utilizaram locos ncSSR para a validação da taxonomia popular feita pelos agricultores, e Duputié et al. (2007) detectaram zonas de hibridização entre a mandioca e seus parentes silvestres. Além disso, marcadores ncSSR vêm sendo amplamente empregados para a estimativa da diversidade genética presente nas variedades de mandioca cultivadas e manejadas tradicionalmente e suas implicações para a

dinâmica evolutiva do cultivo (ALVES-PEREIRA et al., 2011; DUPUTIÉ et al., 2009; PUJOL; DAVID; McKEY, 2005; ROCHA et al., 2008; SARDOS et al., 2008).

Alguns estudos abordaram a questão da organização da diversidade genética de mandioca ao longo do espaço empregando marcadores microssatélites (ELIAS et al., 2004; MÜHLEN; MARTINS; ANDO, 2000; SIQUEIRA et al., 2009). Entretanto, estes estudos foram realizados com amostragens de regiões disjuntas provenientes de bancos de germoplasma, e não foram direcionados para a região amazônica.

2.6 Restriction-site associated DNA sequencing (RAD-seq) e marcadores SNPs

Marcadores do tipo RAD (*Restriction-site associated DNA*) são pequenos fragmentos de DNA adjacentes a sítios de reconhecimento de enzimas de restrição. Estes marcadores foram inicialmente desenvolvidos para a detecção de polimorfismos de sítios únicos (SNPs) por meio de microarranjos (*microarrays*) (MILLER et al., 2007). A metodologia de descobrimento de marcadores SNPs foi então adaptada para as tecnologias de sequenciamento de nova geração, e esta técnica passou a ser denominada RAD-seq ou RAD-tags (BAIRD et al., 2008). Esta técnica permite a seleção de um determinado número de locos RAD com base na escolha de uma enzima de restrição, e o número de locos pode ser aumentado quase que indefinidamente com a utilização de enzimas de restrição adicionais (BAIRD et al., 2008; DAVEY; BLAXTER, 2010). Desta forma, os marcadores SNPs descobertos por meio da técnica de RAD-seq podem ser aplicados na obtenção de dados para abordagens de genômica populacional (DAVEY; BLAXTER, 2010).

O sequenciamento de locos RAD-seq se dá usualmente por meio da plataforma Illumina (BENTLEY et al., 2008). O sequenciamento massivo e paralelo de amostras *multiplex* permite o rápido descobrimento de milhares a milhões de locos RAD-seq e por conseguinte a genotipagem em larga escala de populações com muitas centenas ou milhares de marcadores SNPs.

Para o descobrimento de marcadores SNPs com a técnica de RAD-seq uma amostra de DNA é digerida com uma determinada enzima de restrição que produza fragmentos com extremidades coesivas. Adaptadores P1 específicos (*barcoded adaptors*, que são adaptadores modificados para utilização em equipamentos Illumina) para cada indivíduo da população são então ligados a estes fragmentos. Após a ligação dos adaptadores P1 as amostras são misturadas (*pool*), e submetidas à fragmentação aleatória (*random shearing*), e os fragmentos que possuem entre 300 e 700 pares de bases são selecionados. Adaptadores P2 com uma das extremidades

divergente são adicionados aos fragmentos com e sem os adaptadores P1. Os fragmentos são então amplificados por meio de PCR com iniciadores complementares a P1 e P2. Apenas os fragmentos com adaptadores P1 e P2 serão amplificados, e estes serão sequenciados em plataforma Illumina (DAVEY et al., 2011). Para a detecção dos polimorfismos as sequências geradas podem então ser alinhadas com um genoma de referência, ou na ausência deste, sequências RAD-seq similares podem ser agrupados e tratados como locos candidatos (DAVEY; BLAXTER, 2010). Esta técnica pode ser aplicada tanto em organismos que possuam sequências genômicas de referência, como em organismos não-modelo, para os quais não existem informações. Um esquema do método é ilustrado na Figura 5.

Outros protocolos com variações da técnica original vêm sendo descritos, e diferem basicamente na quantidade e tipos de enzima de restrição utilizados para redução da complexidade do genoma, e nos reagentes e etapas de enriquecimento das bibliotecas de RAD (PURITZ et al., 2014). No *double digest* RAD (ddRAD, PETERSON et al., 2012) duas enzimas de restrição são empregadas na digestão do DNA, e não há necessidade da fragmentação aleatória. No ezRAD (TOONEN et al., 2013) ocorre o emprego de duas enzimas de restrição diferentes mas com a mesma sequência de corte, e os fragmentos são ligados aos adaptadores e *barcodes* com um kit comercial. Já no 2bRAD (WANG et al., 2012) é empregado um tipo de enzima de restrição que gera fragmentos de 36 pares de base, e não existe seleção por tamanho de fragmentos.

O polimorfismo revelado em locos RAD-seq originam-se nas sequências adjacentes aos sítios de restrição, onde são identificados polimorfismos de nucleotídeos únicos (SNPs). Alternativamente, a ocorrência de mutações nos sítios de restrição ao longo do genoma dos indivíduos, causam a presença ou ausência de determinada sequência de RAD-seq nos indivíduos de uma população (BAIRD et al., 2008), podendo então ser também analisados como marcadores dominantes. Em geral, os locos RAD-seq são analisados para o descobrimento de marcadores SNPs.

Figura 5 - Esquema do método de *Restriction-site associated DNA sequencing* (RAD-seq). Amostras de DNA são submetidas à digestão com enzima de restrição e adaptadores P1 contendo diferentes sequências *barcodes* para cada indivíduo (retângulos vermelhos e azuis) são ligados aos fragmentos gerados. Uma etapa de fragmentação aleatória é seguida pela seleção de tamanho dos fragmentos. Adaptadores P2 são ligados aos fragmentos selecionados, e uma PCR seletiva para o enriquecimento das bibliotecas RAD é realizada. As sequências geradas para cada indivíduo podem ser alinhadas a um genoma de referência para o descobrimento de polimorfismos de nucleotídeos únicos (SNPs)

Como o próprio nome sugere, marcadores SNPs (*Single Nucleotide Polymorphisms*) se referem às posições de uma única base ao longo de uma sequência de DNA que é variável entre indivíduos (FREELAND, 2005). Observa-se que a maioria dos marcadores SNPs são bialélicos, possuindo apenas dois estados alternativos, e estima-se que sua taxa de mutação seja da ordem de 10^{-8} a 10^{-9} (FREELAND, 2005). Estes marcadores são os tipos de variantes genéticas mais frequentes e dispersos ao longo do genoma das espécies, estimando-se que correspondam à 90% da variação genética observada em humanos (COLLINS; BROOKS; CHAKRAVARTI, 1998). Em uma análise de múltiplos táxons, incluindo plantas, Brumfield et al. (2003) sugerem que variantes do tipo SNP ocorram a cada 200-500 pares de bases em regiões não-codificadoras, e a cada 500-1000 pares de bases em regiões codificadoras.

As tecnologias de sequenciamento de nova geração permitem o descobrimento de milhares de marcadores SNPs, a um custo cada vez menor, em espécies modelo e não-modelo (TAUTZ; ELLEGREN; WEIGEL, 2010). Este advento das tecnologias de sequenciamento de nova geração, aliado a abundância e ocorrência ao longo de todo o genoma tornaram os polimorfismos SNPs os marcadores de maior apelo para a genômica de populações (WILLING; DREYER; VanOOSTERHOUT, 2012). Ainda há incerteza sobre a quantidade de marcadores SNPs necessários para substituir a informação sobre variação genética revelada por marcadores microssatélites (SEEB et al., 2011). Marcadores ncSSR são hipervariáveis e revelam mais informação por loco do que marcadores SNPs, que usualmente são bialélicos (ZALAPA et al., 2012). Com base na projeção descrita em Zalapa et al. (2012), uma variação equivalente à revelada por 20 marcadores ncSSR com média de 20 alelos seria obtida com 380 locos SNPs. Mesmo assim é provável que, nos estudos de genética de populações, as abordagens baseadas em sequenciamento de nova geração e descobrimento de SNPs tornem-se mais frequentes do que o emprego de marcadores microssatélites (SEEB et al., 2011). Entretanto os marcadores microssatélites devem ainda ser usualmente empregados em abordagens específicas, tais como em análises de paternidade e parentesco (SEEB et al., 2011).

Recentemente marcadores SNPs vêm sendo utilizados em mandioca para diferentes finalidades. Ferguson et al. (2012) validaram 1.190 marcadores SNPs derivados de bibliotecas de ESTs (*Expressed sequence tags*) em 53 variedades de mandioca. Oliveira et al. (2012) avaliaram a utilidade de 390 SNPs derivados de bibliotecas de ESTs relacionados à tolerância a seca para seleção genômica de características de produtividade. Pootakham et al. (2014) utilizaram 2.110 SNPs identificados por RNA-seq de variedades tailandesas de mandioca para a construção de um mapa genético. Peña-Venegas et al. (2014) avaliaram a diversidade genômica de 173 variedades colombianas de mandioca com base em 93 marcadores SNPs.

Oliveira et al. (2014) avaliaram os padrões de estrutura genética de 1.280 acessos do banco brasileiro de germoplasma de mandioca com base em 402 marcadores SNPs.

O emprego de marcadores ncSSR, cpSSR e SNPs no estudo da diversidade genética e genômica de mandioca cultivada tradicionalmente ao longo dos principais rios da bacia amazônica brasileira permite que seja feita uma avaliação mais clara e robusta dos padrões de dispersão das variedades de mandioca mansas e bravas na Amazônia. Além disso, o emprego de marcadores SNPs possibilita a detecção de sequências ao longo do genoma que apresentem desvios do comportamento esperado para marcadores neutros (*outliers*). Tais locos podem estar associados à assinaturas de seleção positiva, balanceadora ou diversificadora, por exemplo (DAVEY; BLAXTER, 2010). Desta forma, a detecção de locos *outliers* permite a identificação de sequências do genoma que podem estar associadas à diferenciação entre variedades de mandioca mansas e bravas, bem como associadas a possíveis padrões distintos de dispersão. Informações mais detalhadas sobre a dispersão e difusão da mandioca na região amazônica, bem como informações inéditas sobre a organização da diversidade genômica, têm importância prática. Regiões em que o cultivo apresenta alta diversidade genética poderão ser identificadas, o que por sua vez poderá servir para orientar futuras prospecções e esforços de conservação dos recursos genéticos da mandioca, tanto *in situ* como *ex situ*, e que também poderiam direcionar estratégias para o melhoramento do cultivo.

2.7 Genômica populacional

Em uma definição geral, a genômica de populações é o estudo de vários locos ou regiões ao longo de todo o genoma objetivando o melhor entendimento do papel dos processos evolutivos que influenciam a variação ao longo do genoma e das populações (LUIKART et al., 2003). A genômica de populações avalia vários locos do genoma para identificar e separar efeitos loco-específicos (como seleção, mutação e recombinação) daqueles que se refletem ao longo de regiões genômicas mais extensas (como deriva genética, *bottlenecks*, fluxo gênico e endogamia), para o melhor entendimento de processos microevolutivos (ALLENDORF; HOHENLOHE; LUIKART, 2010; BLACK et al., 2001). Desta forma, os dois princípios básicos da genômica de populações são que os locos neutros do genoma serão afetados de forma similar pela história demográfica e evolutiva das populações, enquanto que locos que estão sob seleção frequentemente apresentarão comportamento diferenciado (*outliers*) com relação à variação genética (LUIKART et al., 2003).

Uma abordagem genômica se inicia com a amostragem dos indivíduos de uma ou mais populações, mas não necessariamente requer uma definição *a priori* entre os limites das populações. Após a amostragem, o próximo passo é o descobrimento do polimorfismo em dezenas a centenas de locos (LUIKART et al., 2003). Com o advento das tecnologias de sequenciamento de nova geração é possível o descobrimento de milhares de marcadores ao longo do genoma de interesse, mesmo para as populações de espécies em que se tem pouca ou nenhuma informação genética (DAVEY et al., 2011). O terceiro passo é a detecção de locos *outliers*, sendo talvez o mais importante, já que a maioria dos princípios da genética de populações requer o uso de marcadores neutros. Os locos *outliers* estão hipoteticamente sob seleção, e podem ser detectados, por exemplo, por apresentarem valores extremamente altos ou baixos de F_{ST} entre populações, por excessos ou déficits de heterozigosidade, ou desvios muito grandes do Equilíbrio de Hardy-Weinberg em um dado loco. O quarto passo da abordagem genômica diz respeito ao uso dos locos neutros para a inferência da história demográfica de uma população. Do mesmo modo, os locos *outliers* podem ser utilizados, por exemplo, em estudos de adaptação local (OUBORG et al., 2010).

Segundo Luikart et al. (2003), o entendimento da evolução adaptativa, bem como a melhora da inferência dos parâmetros populacionais e da reconstrução da história evolutiva de populações podem ser consideradas as maiores influências da genômica de populações para a genética de populações. Isto porque devido ao grande número de marcadores utilizados a genômica de populações permite a identificação dos locos *outliers* e o aumento da cobertura genômica, o que por sua vez aumenta a acurácia e precisão das análises. As abordagens de genômica de populações permitem a avaliação simultânea da história filogenética e demográfica de uma população, por meio de locos seletivamente neutros, e a identificação de regiões específicas do genoma que respondem à seleção local (ALLENDORF; HOHENLOHE; LUIKART, 2010). Desta forma, elas podem aumentar a sensibilidade de muitas investigações moleculares sobre a ecologia e genética de populações, e de genética da conservação.

Exemplos clássicos de estudos baseados em genômica populacional com diferentes enfoques estão nos trabalhos pioneiros de evolução adaptativa do peixe esgana-gato *Gasterosteus aculeatus* (HOHENLOHE et al., 2010), e de filogeografia do mosquito fitófago *Wyeomyia smithii* (EMERSON et al., 2010). Com base em aproximadamente 45.000 marcadores SNPs, Hohenlohe et al. (2010) identificaram regiões do genoma com assinaturas de seleção divergente entre populações marinhas e lacustres de *G. aculeatus*. Emerson et al. (2010) identificaram 3.741 marcadores SNPs e encontraram padrões concordantes de diversificação e dispersão das populações de *W. smithii* com a expansão das populações da

planta em que ele ocorre após o último período glacial. Este estudo também foi importante para demonstrar que abordagens genômicas podem ser aplicadas também em espécies não-modelo, as quais possuem pouca informação sobre o genoma. Exemplos de estudos genômicos com plantas foram realizados em duas espécies de *Populus* (STÖLTING et al., 2013), e em berinjela *Solanum melongena* (BARCHI et al., 2011). Stölting et al. (2013) usaram mais de 38.000 marcadores SNPs para avaliar os diferentes níveis de divergência ao longo do genoma de *Populus tremula* e *P. alba*. Para berinjela, Barchi et al. (2011) descobriram cerca de 10.000 marcadores SNPs, dos quais 384 foram arranjados em um painel para a caracterização de germoplasma. Para a identificação de marcadores SNPs, todos os estudos supracitados usaram a técnica de RAD-seq, apresentada um pouco mais detalhadamente no tópico 2.6.

2.8 Filogeografia

A filogeografia estuda os princípios e processos que governam a distribuição geográfica das linhagens genealógicas, especialmente aquelas dentro e entre espécies proximamente relacionadas (AVISE, 2004). Este campo de estudo está focado explicitamente nos componentes filogenéticos ou históricos da estruturação genética populacional (incluindo aqueles que podem ser influenciados pela deriva genética, fluxo gênico, seleção natural, dentre outros fatores evolutivos) (AVISE, 2009).

A introdução de marcadores baseados no DNA mitocondrial (mtDNA) em estudos de genética de populações no fim dos anos 1970 proporcionou a criação de uma perspectiva genealógica sobre a estrutura populacional das espécies. Em comparação com o mtDNA de plantas, o DNA do cloroplasto (cpDNA) apresenta menos rearranjos e maior taxa de substituição de nucleotídeos (AVISE, 2004). Para a maioria dos organismos superiores o genoma mitocondrial e cloroplastidial são transmitidos de maneira uniparental (para a maioria das angiospermas o cpDNA é herdado maternalmente). Uma vez que sequências do cpDNA evoluem rapidamente e a recombinação raramente ocorre, elas fornecem dados haplotípicos que podem ser ordenados filogeneticamente dentro de espécies. Os resultados são filogenias intraespecíficas interpretadas como o componente materno em linhagens de descendência entre indivíduos. Por esta última razão, o cpDNA tem sido especialmente útil para análises filogeográficas de espécies de plantas (AVISE, 2009).

Análises filogeográficas em plantas domesticadas têm sido importantes para a elucidação dos eventos de domesticação e dos padrões de dispersão dos cultivos a partir dos seus centros de origem. Com base na análise de sequências do espaçador *trnG-trnS* do

cloroplasto em árvores cultivadas e silvestres de seriguela (*Spondias purpurea*) e mais duas espécies do gênero, Miller e Schaal (2005) identificaram prováveis eventos múltiplos de domesticação desta fruteira Mesoamericana. Os padrões de diversidade nucleotídica e estruturação haplotípica de cinco locos nucleares foram usados para sugerir um único evento de domesticação na Mesoamérica para o feijão comum (*Phaseolus vulgaris*, BITOCCHI et al., 2012). Avaliando a variação de marcadores microssatélites do cloroplasto e do núcleo em acessos de batata-doce (*Ipomoea batatas*), Roullier et al. (2011) sugeriram pelo menos dois eventos de domesticação independentes na América Central e Caribe, e na porção noroeste da América do Sul. Foi também a partir de uma abordagem filogeográfica que Olsen e Schaal (1999) demonstraram que a forma cultivada de mandioca (*Manihot esculenta* ssp. *esculenta*) foi domesticada a partir de populações de *Manihot esculenta* ssp. *flabellifolia*. Neste estudo os autores utilizaram a variação no gene nuclear de cópia única gliceraldeído 3-fosfato desidrogenase (*G3pdh*) e apontaram a região sudoeste da bacia amazônica como a provável região onde a mandioca foi domesticada.

Análises genéticas de 494 variedades brasileiras de mandioca mansas e bravas provenientes de 11 regiões geográficas (cinco delas amazônicas) em cinco biomas brasileiros revelaram diferentes padrões de estruturação genética dentro das variedades mansas e bravas (informação verbal)¹. Entretanto, foram utilizados apenas marcadores microssatélites nucleares, e não foi realizada uma amostragem sistemática, sendo que as variedades provieram tanto de prospecções, como de bancos de germoplasma. Além disso, falta no estudo uma amostragem que inclua a região de Rondônia para que se tenha uma visão mais clara de como a dispersão diferencial das variedades mansas e bravas de mandioca teria ocorrido a partir do centro de origem do cultivo.

¹ Gilda Santos Mühlen. Universidade Federal de Rondônia (UNIR).

3 MATERIAL E MÉTODOS

3.1 Material Vegetal

Foram visitados um total de 44 municípios, em quatro estados brasileiros, totalizando 596 indivíduos avaliados, sendo 325 variedades de mandioca brava, 226 variedades mansas, 17 não designadas e 28 indivíduos *M. esculenta* ssp. *flabellifolia* (a partir de agora designados como silvestres) (Figura 6 e Tabela 1). Amostras de folhas de um indivíduo de cada variedade presente em roçados identificados em cada município foram coletadas e acondicionadas em sacos plásticos com sílica gel para desidratação e conservação até o procedimento de extração do DNA. Desde a chegada ao laboratório as amostras estão sendo mantidas a - 20°C em freezers. Os roçados visitados tiveram sua localização geográfica registrada usando o Sistema de Posicionamento Global (GPS). As amostragens realizadas foram sempre e exclusivamente de folhas e fotos, para poder caracterizar as prospecções como pesquisa científica e evitar protocolos demorados para a obtenção de uma autorização junto ao Conselho de Gestão do Patrimônio Genético (CGEN-MMA).

Figura 6 - Mapa ilustrando as localidades (pontos) onde variedades de mandioca mansas e bravas foram coletadas junto às comunidades de agricultores tradicionais ao longo dos rios Madeira, Negro, Branco, Solimões e Amazonas. Também foram amostradas variedades do sul de Rondônia (rio Guaporé) e nordeste do Pará

Tabela 1 - Locais de coleta (Estados e Municípios) com os respectivos números de variedades de mandiocas mansas, bravas, não designadas (ND) e indivíduos silvestres amostrados. O georreferenciamento indica a sede dos municípios visitados

(continua)

Estado	Município	Localização	Mandioca				Total
			Brava	Mansa	ND	Silvestre	
Amazonas			246	87	3	-	336
	Alvarães	03°13'15" S 64°48'15" O	13	2	-	-	15
	Autazes	03°34'48" S 59°07'51" O	16	6	-	-	22
	Barcelos	00°58'30" S 62°55'26" O	43	10	-	-	53
	Borba	04°23'16" S 59°35'38" O	12	-	-	-	12
	Careiro Castanho	03°46'04" S 60°22'08" O	7	-	-	-	7
	Codajás	03°50'13" S 62°03'25" O	1	1	-	-	2
	Fonte Boa	02°30'50" S 66°05'30" O	2	2	-	-	4
	Humaitá	07°30'22" S 63°01'15" O	15	12	3	-	30
	Itapiranga	02°44'56" S 58°01'19" O	-	3	-	-	3
	Manaquiri	03°25'41" S 60°27'34" O	6	5	-	-	11
	Manicoré	05°48'32" S 61°18'00" O	66	8	-	-	74
	Novo Aripuanã	05°08'00" S 60°22'30" O	-	6	-	-	6
	Parintins	02°37'40" S 56°44'09" O	4	13	-	-	17
	Santa Isabel	00°24'50" S 65°01'08" O	17	6	-	-	23
	Santo Antônio	03°06'07" S 67°56'24" O	3	-	-	-	3
	São Gabriel	00°07'48" S 67°05'20" O	23	3	-	-	26
	São Paulo	03°22'40" S 68°52'22" O	4	2	-	-	6
	Silves	02°50'20" S 58°12'33" O	8	8	-	-	16
	Tabatinga	04°15'09" S 69°56'17" O	3	-	-	-	3
	Tefé	03°21'14" S 64°42'38" O	3	-	-	-	3
Pará			23	32	-	-	55
	Almeirim	01°31'22" S 52°34'55" O	-	3	-	-	3
	Bom Jesus	05°03'03" S 48°36'32" O	1	2	-	-	3
	Marabá	05°22'08" S 49°07'04" O	4	-	-	-	4
	Monte Alegre	01°59'56" S 54°04'58" O	5	4	-	-	9
	Oriximiná	01°45'57" S 55°51'57" O	1	4	-	-	5
	Paragominas	02°59'42" S 47°21'10" O	-	3	-	-	3
	Rondon do Pará	04°46'33" S 48°04'01" O	-	3	-	-	3
	Santa Maria	01°21'07" S 47°34'33" O	4	-	-	-	4
	Santarém	02°26'34" S 54°42'28" O	8	8	-	-	16
	São Miguel	01°37'37" S 47°28'58" O	-	5	-	-	5
Rondônia			20	38	11	28	97
	Cabixi	13°13'52" S 60°33'15" O	-	8	4	-	12
	Guajará Mirim	10°46'58" S 65°20'22" O	16	23	4	2	45
	Nova Mamoré	10°24'07" S 65°19'36" O	2	2	-	1	5
	Pimenteiras D'oeste	13°28'57" S 61°02'48" O	2	5	3	-	10
	Rolim de Moura	11°48'13" S 61°48'12" O	-	-	-	25	25
Roraima			36	69	3	-	108
	Alto Alegre	02°58'48" N 61°17'31" O	-	3	-	-	3

Tabela 1 - Locais de coleta (Estados e Municípios) com os respectivos números de variedades de mandiocas mansas, bravas, não designadas (ND) e indivíduos silvestres amostrados. O georreferenciamento indica a sede dos municípios visitados

(conclusão)

Estado	Município	Localização	Mandioca				Total
			Brava	Mansa	ND	Silvestre	
Roraima			36	69	3	-	108
	Bonfim	03°21'36" N 59°49'58" O	11	8	-	-	19
	Cantá	02°36'36" N 60°35'49" O	1	1	-	-	2
	Caracarái	01°48'57" N 61°07'40" O	13	15	-	-	28
	Mucajái	02°25'48" N 60°54'00" O	4	13	-	-	17
	Normandia	03°52'51" N 59°37'22" O	3	6	-	-	9
	Rorainópolis	00°56'45" N 60°25'04" O	4	7	3	-	14
	São João da Baliza	00°57'03" N 59°54'39" O	-	8	-	-	8
	São Luís do Anauá	01°00'18" N 60°09'32" O	-	8	-	-	8
Total			325	226	17	28	596

As amostragens foram realizadas com o apoio financeiro e logístico de vários projetos. No âmbito do projeto CT-Amazônia/CNPq 575588/08-0, coordenado pelo pesquisador Charles R. Clement (INPA), foram realizadas quatro prospecções para a coleta de amostras de variedades mansas e bravas de mandioca cultivadas tradicionalmente ao longo dos principais eixos fluviais da bacia amazônica (rios Madeira, Negro/Branco, Solimões e Amazonas).

Para o incremento da amostragem foram realizadas amostragens nos municípios de Tefé e Alvarães (na calha do rio Solimões), em São Gabriel da Cachoeira (na calha do rio Negro), em Novo Aripuanã (na calha do rio Madeira), e ainda em Silves e Itapiranga (na calha do rio Amazonas), todos no estado do Amazonas. Estas atividades contaram com o apoio dos projetos CNPq Universal 473422/2012-3, FAPEAM Universal 3137/2012, ambos liderados pelo pesquisador Charles R. Clement (INPA).

Também foram visitados os municípios de Pimenteiras D' oeste, Cabixi e Rolim de Moura, no estado de Rondônia, visando a amostragem de variedades de mandiocas cultivadas em regiões próximas ao centro de origem e domesticação do cultivo. No município de Rolim de Moura foram amostrados 25 indivíduos silvestres visando uma análise comparativa dos padrões de diversidade genética e agregação de consistência às inferências acerca da história evolutiva do cultivo. Amostragens oportunistas também foram realizadas nos municípios de Marabá, Rondon do Pará, Paragominas, São Miguel do Guamá, Bom Jesus e Santa Maria, todos no estado do Pará. A inclusão de variedades de mandioca destes municípios contribuiu para a comparação dos níveis de diversidade genética manejados em variedades de mandioca cultivadas em regiões mais afastadas dos principais rios amazônicos. Estas atividades foram realizadas em parceria com a pesquisadora Elizabeth Ann Veasey, (ESALQ/USP), com

recursos do projeto FAPESP 2012/08307-5. Todas as atividades de amostragem foram apoiadas também com recursos do projeto FAPESP 2013/00003-0 coordenado pela pesquisadora Maria Imaculada Zucchi (APTA-Pólo Centro-Sul).

3.2 Extração e quantificação de DNA

O DNA genômico total dos indivíduos amostrados foi extraído a partir de aproximadamente 50 mg de amostras de folhas utilizando o protocolo de extração com o detergente catiônico CTAB 2% (DOYLE; DOYLE, 1990). A quantificação do DNA genômico foi realizada por meio de eletroforese em gel de agarose 0,9% (p/v) corado com SYBR Safe DNA Gel Stain (Invitrogen), por comparação visual com amostras de DNA do fago lambda de concentrações conhecidas.

3.3 Amplificação e detecção do polimorfismo de marcadores cpSSR

Foram testados 20 locos microssatélites cloroplastidiais (cpSSR) universais para dicotiledôneas descritos para tabaco: ccmp1 a ccmp10 (WEISING; GARDNER, 1999), além dos locos ccSSR1, 2, 4, 7, 9, 11, 16, 19, 20 e 22 (CHUNG; STAUB, 2003). Os testes de amplificação dos locos foram realizados por meio de PCR com um volume final de 10 µL, contendo 10 ng de DNA genômico, 1X de tampão (10 mM Tris-HCl; pH 8,3; 50 mM KCl; 1,5 mM MgCl₂), 2,5 ng de BSA, 2,5mM de MgCl₂, 200 µM de cada um dos dNTPs, 1,6 pmols do iniciador *forward*, 2 pmols do iniciador *reverse*, 1,5 pmols do iniciador M13(-29) e 1U de *Taq* DNA polimerase. As amplificações foram realizadas em termociclador ProFlex (Applied Biosystems) com as seguintes condições: 94°C por 4 minutos; 30 ciclos a 94°C por 1 minuto, 58°C por 1 minuto e 72°C por 1 minuto; seguidos de 10 ciclos a 94°C por 40 segundos, 53°C por 40 segundos e 72°C por 40 segundos, além de um passo final a 72°C por 10 minutos para extensão complementar dos fragmentos. A qualidade das amplificações dos locos cpSSR foi verificada visualmente por meio de eletroforese em gel de agarose (2% p/v) com os produtos da PCR corados com SYBR Safe (Invitrogen).

Para a detecção do polimorfismo dos locos cpSSR um dos iniciadores foi marcado com um fluoróforo específico (IRDye-700 ou IRDye-800). Os produtos obtidos por meio de PCR foram analisados conjuntamente em sistemas *biplex* (com dois locos marcados com fluorescências diferentes e analisados em conjunto) no equipamento 4300 DNA Analyzer (LI-COR). Os produtos das reações de PCR foram submetidos à uma eletroforese em gel de

acrilamida (6,5%) e detectados automaticamente por meio de um sensor de raio *laser*. O processamento e determinação dos genótipos das amostras foi realizado utilizando-se o sistema SAGA^{GT} (LI-COR), e o tamanho dos fragmentos detectados foi feito com auxílio dos padrões de peso molecular 50-350 bp IRDye 700 *Sizing Standard* e 50-350 bp IRDye 800 *Sizing Standard* (LI-COR) (Figura 7).

Figura 7 - Esquema das etapas de amplificação e detecção de polimorfismo com marcadores microssatélites cloroplastidiais (cpSSR). A) amplificação de locos cpSSR por meio de PCR e detecção de polimorfismos em sequenciador Li-Cor 4300 *Analyzer*; B) os genótipos dos indivíduos foram determinados por meio da análise de géis virtuais, e a matriz de genótipos resultante foi utilizada para as análises estatísticas descritas no tópico 3.6.1

3.4 Amplificação e detecção do polimorfismo de marcadores ncSSR

Foram testados 15 locos microssatélites do genoma nuclear (ncSSR): GA12, GA21, GA126, GA127, GA131, GA134, GA136, GA140, GAGG5 (CHAVARRIAGA-AGUIRRE et al., 1998); e os locos SSRY13, SSRY20, SSRY32, SSRY70, SSRY89 e SSRY164 (MBA et al., 2001). Estes locos foram amplificados por meio de PCR em termociclador ProFlex (Applied Biosystems) como descrito em Alves-Pereira et al. (2011).

A detecção dos polimorfismos realizada em sequenciador automático foi executada junto ao Centro de Biotecnologia Molecular e Engenharia Genética do Instituto de Biologia da Universidade Estadual de Campinas (CBMEG IB-Unicamp). Para a detecção do polimorfismo dos locos ncSSR os iniciadores foram marcados com um fluoróforo específico (6-FAM, NED

ou HEX). As amostras foram amplificadas via PCR e os produtos foram analisados conjuntamente em sistemas *multiplex*. Aos produtos das reações de PCR foi adicionado uma mistura de formamida Hi-Di (Applied Biosystems) e um marcador de peso molecular (GeneScan 500 ROX - *Size Standard*, Applied Biosystems). Os fragmentos SSR foram então submetidos à separação por eletroforese capilar em sequenciador automático ABI3500xl (Applied Biosystems). A análise e determinação dos tamanhos dos fragmentos SSR foi realizada com auxílio do programa *Gene Mapper* v.4.0 (Applied Biosystems), o qual gerou um relatório com os genótipos de cada indivíduo (Figura 8).

Figura 8 - Esquema das etapas de amplificação e detecção de polimorfismo com marcadores microssatélites nucleares (ncSSR). A) amplificação de locos ncSSR por meio de PCR e detecção de polimorfismos em sequenciador ABI 3500xl; B) os genótipos dos indivíduos foram determinados por meio da análise de eletroferogramas, e a matriz de genótipos resultante foi utilizada para as análises estatísticas descritas no tópico 3.6.2

3.5 Detecção do polimorfismo de marcadores SNPs com a técnica de RAD-seq

Devido ao alto custo da técnica para o descobrimento de marcadores RAD-seq, foram selecionados 159 indivíduos de modo a representar cada grupo de variedades cultivadas para cada eixo fluvial amostrado, sendo também inclusos indivíduos silvestres (Figura 9). Os SNPs foram detectados em 71 indivíduos de variedades bravas, 69 mansas e 19 silvestres. Quando considerados os rios, foram sequenciados 37 indivíduos do rio Madeira (18 mansas, 19 bravas), 40 dos rios Negro e Branco (17 mansas, 23 bravas), 33 do rio Solimões (16 mansas, 17 bravas) e 30 do rio Amazonas (18 mansas, 12 bravas).

Figura 9 - Mapa ilustrando as localidades (pontos) de onde as variedades de mandioca mansas e bravas incluídas nas análises de diversidade e estrutura genômica foram coletadas junto às comunidades de agricultores tradicionais ao longo dos rios Madeira, Negro, Branco, Solimões e Amazonas. Também foram amostrados indivíduos silvestres no estado de Rondônia

Como recomendação da empresa Floragenex (prestadora do serviço de descobrimento de SNPs pela técnica de RAD-seq), e visando a obtenção de material de alta pureza e em grande quantidade, o DNA dos indivíduos selecionados foi extraído com kits de extração DNeasy Plant Mini Kit (Quiagen) ou Nucleospin Plant II (Marcherey-Nagel) de acordo com os protocolos sugeridos pelos fabricantes.

Os marcadores SNPs foram descobertos utilizando-se aproximadamente 1µg de DNA de cada indivíduo marcados com uma sequência única de seis nucleotídeos (*barcode*), e as amostras foram submetidas a um equipamento Illumina (*Genome Analyzer Illumina II*), como descrito por Hohenlohe et al. (2011). Resumidamente, sequências de adaptador e *barcodes* foram ligadas aos fragmentos do DNA genômico total digeridos com a enzima de restrição *Sbf*I. Os fragmentos associados aos sítios de restrição desta enzima foram então submetidos ao sequenciamento na plataforma *Illumina Genome Analyzer IIx* com a configuração *single-end* e 91pb (Figura 10). As sequências geradas foram então separadas por indivíduo e tiveram suas sequências *barcodes* removidas, resultando em fragmentos RAD de cerca de 80 pb. Estes ensaios foram realizados pela empresa Floragenex, Inc. situada no estado de Oregon, nos Estados Unidos, e utilizaram verba destinada à prestação de serviços de terceiros no exterior previstos para este no projeto FAPESP 2013/00003-0.

Figura 10 - Esquema das etapas de descobrimento de polimorfismos de nucleotídeos únicos (SNPs) por meio da técnica de *Restriction-site associated DNA* (RAD-seq). A) e B) as bibliotecas RAD são preparadas e submetidas ao sequenciamento na plataforma *Illumina Genome Analyzer IIx*; C) As sequências obtidas foram separadas por indivíduos ("demultiplexagem"), e alinhadas a um genoma de referência; D) Marcadores SNPs foram descobertos nos locos RAD identificados pelo alinhamento, e a matriz de genótipos resultante foi utilizada para as análises estatísticas descritas no tópico 3.6.3

3.6 Análises Genéticas

3.6.1 cpSSR

A estimação do número de alelos (A), o índice de informação correspondente ao índice de Shannon-Weaver (I , BROWN; WEIR, 1983) e a diversidade gênica de Nei (1987, h) para locos individuais foram realizados com o programa *GenAlEx 6.5* (PEAKALL; SMOUSE, 2012). Combinações multilocos únicas de alelos dos cpSSRs foram tratadas como haplótipos diferentes. Para a detecção dos haplótipos e para a análise da diversidade haplotípica foram estimados o número total (N_H) e efetivo (N_{eH}) de haplótipos, diversidade haplotípica (H_{Ehap})

também com o programa *GenAlEx* 6.5 (PEAKALL; SMOUSE, 2012). A diversidade nucleotídica (π) foi estimada com o programa *Arlequin* v.3.5 (EXCOFFIER; LISCHER, 2010).

Para avaliar a estrutura genética existente entre as variedades de mandioca e seus locais de coleta foram realizadas Análises Bayesianas com o programa *Structure* v.2.3.3 (PRITCHARD; STEPHENS; DONNELLY, 2000). Foram realizadas 10 simulações para cada K número de agrupamentos (com K variando de 1 a 20), considerando o modelo de frequências alélicas correlacionadas e o modelo *no admixture* de ancestralidade, com 500.000 iterações da cadeia de Markov Monte Carlo (MCMC) após um período inicial de *burnin* de 200.000 iterações. O número K mais provável de agrupamentos foi determinado pelo método de Evanno, Regnaut e Goudet (2005) baseado em ΔK , com auxílio do aplicativo *on-line Structure Harvester* (EARL; VonHOLDT, 2012). O programa *Clumpp* v. 1.1.2b (JAKOBSSON; ROSENBERG, 2007) foi utilizado para sumarizar os resultados obtidos por cada uma das 10 simulações realizadas para o número K de agrupamentos mais provável. A representação em gráficos de barra dos resultados obtidos foi realizada com o programa *Distruct* (ROSENBERG, 2004).

As análises Bayesianas implementadas no programa *Structure* possuem fortes pressupostos genéticos, tais como aderência dos grupos ao Equilíbrio de Hardy-Weinberg e minimização do desequilíbrio de ligação. Desta forma, a estrutura genética observada com base na variação haplotípica de indivíduos de mandioca foi avaliada também por meio de métodos estatísticos multivariados. Foram realizadas Análises Discriminantes de Componentes Principais (DAPC, JOMBART; DEVILLARD; BALLOUX, 2010) com o pacote da plataforma R (R DEVELOPMENT CORE TEAM, 2003) *Adegenet* v.1.3.1 (JOMBART; AHMED, 2011). Resumidamente, a DAPC busca por combinações lineares de alelos que representem o máximo de variação possível entre grupos, ao mesmo tempo em que minimiza as diferenças dentro dos grupos (ROULLIER et al., 2013). Com base nas funções discriminantes o método gera para cada indivíduo uma probabilidade de adesão aos diferentes grupos (JOMBART; DEVILLARD; BALLOUX, 2010). Tais probabilidades que podem ser interpretados de forma similar às probabilidades de ancestralidade geradas pelas análises Bayesianas implementadas pelo programa *Structure*.

A distribuição hierárquica da variação haplotípica dentro e entre grupos de variedades foi avaliada com Análises de Variância Molecular (AMOVA, MICHALAKIS; EXCOFFIER, 1996) com o programa *Arlequin* v.3.5 (EXCOFFIER; LISCHER, 2010). A significância das medidas de divergência foi testada com 20.000 permutações. Para a determinação da genealogia entre as linhagens matrilineares dos indivíduos de mandioca o relacionamento entre os haplótipos foi avaliado com base na construção de uma rede de haplótipos usando o método

Median-Joining (BANDELT; FORSTER; RÖHL, 1999) com o programa *Network* (FORSTER; BANDELT; RÖHL, 2000). As AMOVAs e rede de haplótipos levaram em consideração apenas os indivíduos que não apresentaram nenhum alelo faltante.

3.6.2 ncSSR

As frequências alélicas para cada loco ncSSR, as estimativas dos parâmetros de diversidade genética [número de alelos (A), riqueza alélica (A_R), número de alelos privados (A_p), número de genótipos multilocos (MLGs), as heterozigosidades observada (H_o) e esperada (H_E)], além dos coeficientes de endogamia (f) (e sua significância baseada em 1.000 réplicas *bootstrap*) foram calculadas para cada loco e para os grupos de indivíduos utilizando os pacotes da plataforma R (R DEVELOPMENT CORE TEAM, 2003) *hierfstat* (GOUDET, 2005) e *poppr* (KAMVAR; TABIMA; GRÜNWALD, 2014).

Análises espaciais da distribuição da diversidade alélica foram realizadas com o programa DIVA-GIS (www.diva-gis.org). Considerando os 596 indivíduos avaliados com 14 locos microssatélites, o total de alelos georreferenciados analisados foi de 16.688 (todos os indivíduos), 9.100 (variedades bravas) e 6.328 (variedades mansas). Foram gerados *grids* com base no número de alelos diferentes presente em células com diâmetro de 10 minutos (correspondendo à aproximadamente 18 Km da área de estudo). Para considerar a diversidade alélica presente nas células adjacentes, foi usada a opção de *circular neighborhood* com diâmetro de 0,45045 graus (correspondendo à aproximadamente 50 Km da área de estudo). Com esta opção, cada célula recebe o valor de diversidade encontrado dentro de um círculo, com o diâmetro em questão, centrado nas células base de 10 minutos (SCHELDEMAN; van ZONNEVELD, 2010). Com isso espera-se gerar *grids* que contenham células com boa resolução e que apresentem menor probabilidade de perda do sinal dos padrões espaciais gerais (SCHELDEMAN; van ZONNEVELD, 2010). Uma vez que o número de alelos é dependente do tamanho amostral existe a tendência de se detectar mais alelos em amostras com maior número de indivíduos (ALLENDORF; LUIKART; AITKEN, 2013). Para melhorar o poder de comparação entre áreas com números amostrais diferentes foi aplicada a rarefação para a estimação da riqueza alélica (ELMOUSADIK; PETIT, 1996). Esta estimativa de diversidade alélica leva em consideração o tamanho amostral. Desta forma, foram realizadas análises espaciais da distribuição de riquezas alélicas considerando para a construção de *grids* células de 10 minutos que contivessem pelo menos 56 observações. Devido a limitação do programa, esta análise não pôde ser feita aplicando o método de *circular neighborhood*.

Para avaliar a estrutura genética foram realizadas análises Bayesianas com o programa *Structure* v.2.3.3 (PRITCHARD; STEPHENS; DONNELLY, 2000), da mesma forma como descrito para os marcadores cpSSR. Foram três realizadas análises: (i) com todos os 596 indivíduos amostrados, (ii) com as 325 mandiocas bravas, e (iii) com as 226 mandiocas mansas. Para a representação geográfica dos agrupamentos identificados, a interpolação espacial das probabilidades de ancestralidade (q) foi realizada pelo método Kriging, para cada um dos agrupamentos identificados pela análise Bayesiana de estruturação genética. Foi utilizada uma função disponibilizada por Oliver François (*Université Joseph Fourier, Institut National Polytechnique de Grenoble*) na plataforma R (R DEVELOPMENT CORE TEAM, 2003). A estrutura genética dos grupos de variedades de mandioca e seus locais de coleta também foi avaliada com métodos de estatística multivariada, por meio de DAPCs realizadas com o pacote da plataforma R (R DEVELOPMENT CORE TEAM, 2003) *Adegenet* v.1.3.1 (JOMBART; AHMED, 2011).

A distribuição hierárquica da variação genética entre e dentro dos grupos de indivíduos foi avaliada com base nas médias de Análises de Variância Molecular (AMOVA, MICHALAKIS; EXCOFFIER, 1996) efetuadas para cada loco (*locus-by-locus* AMOVA) com o programa *Arlequin* v.3.5 (EXCOFFIER; LISCHER, 2010). O nível de significância foi obtido com base em 20.000 permutações.

O relacionamento dos grupos de indivíduos foi avaliado por meio da construção de dendrogramas baseados em distâncias genéticas de Cavalli-Sforza e Edwards (1967), calculadas com o programa *MSA* v.4.05 (DIERINGER; SCHLÖTTERER, 2003). Os dendrogramas foram construídos com o método de agrupamento *Neighbor-Joining* (SAITOU; NEI, 1987), com 1.000 réplicas *bootstrap*, com auxílio do programa *Phylip* v.3.6 (FELSENSTEIN, 2005). As árvores finais foram editadas com o programa *FigTree* v.1.4.1 (<http://tree.bio.ed.ac.uk/software/figtree/>).

3.6.3 RAD-seq

Para o processamento dos fragmentos RAD sequenciados e para identificação de polimorfismos SNPs foi utilizado como referência o genoma parcial da mandioca *Manihot esculenta* v4.1 (PROCHNIK et al., 2012) disponível no banco de dados *Phytozome* v.10.2 (<http://www.phytozome.net/cassava.php>). A identificação de regiões candidatas a locos SNPs foi baseada nos algoritmos dos programas *Bowtie* v. 0.12.8 (LANGMEAD et al., 2009) e *SAMtools* v. 0.1.16 (LI; HANDSAKER; WYSOKER, 2009) e *scripts* personalizados da empresa

FloraGenex. O alinhamento dos fragmentos RAD ao genoma de referência com o programa *Bowtie* levou em consideração as informações sobre a qualidade dos fragmentos sequenciados, permitindo que houvesse até três nucleotídeos diferentes entre cada sequência e a referência. Foram ignorados os fragmentos que alinharam com mais de uma região do genoma ou que não alinharam com região alguma. A tabulação dos locos SNPs candidatos encontrados em todos os indivíduos foi realizada com auxílio do programa *SAMtools* (usando o módulo “*mpileup*”). A identificação dos locos SNPs candidatos foi realizada seguindo um critério com três parâmetros principais: (i) presença da variante em no mínimo 90% dos indivíduos; (ii) o valor médio de qualidade de no mínimo 20 (escala *Phred*) para as sequências que continham as variantes e; (iii) a profundidade mínima de 15X de sequenciamento de cada variante. As informações sobre os locos SNPs e as variantes genômicas encontradas foram então codificadas e armazenadas em arquivos *Variant Calling Format* (VCF, *Vcard files*).

As estimativas de diversidade genômica para cada loco SNP [número total de alelos (A), riqueza alélica média por loco (A_R), heterozigosidades observada (H_O) e esperada (H_E)], além dos coeficientes de endogamia (f) (e sua significância baseada em 1.000 réplicas *bootstrap*) foram calculadas para cada grupo de variedades utilizando o pacote da plataforma R (R DEVELOPMENT CORE TEAM, 2003) *hierfstat* (GOUDET, 2005).

A distribuição hierárquica da variação genômica entre e dentro dos grupos de indivíduos foi avaliada com base nas médias de Análises de Variância Molecular (AMOVA, MICHALAKIS; EXCOFFIER, 1996) efetuadas para cada loco (*locus-by-locus* AMOVA) com o programa *Arlequin* v.3.5 (EXCOFFIER; LISCHER, 2010). O nível de significância foi obtido com base em 20.000 permutações. O grau de diferenciação genética (F_{ST}) par a par entre grupos de indivíduos e suas significâncias com base em 1.000 réplicas *bootstrap* foram calculados com o pacote da plataforma R (R DEVELOPMENT CORE TEAM, 2003) *hierfstat* (GOUDET, 2005).

Para a identificação de locos que possivelmente apresentam assinaturas de seleção (locos *outliers*) foi utilizado o programa *fdist2* (BEAUMONT; NICHOLS, 1996) implementado com o aplicativo *Lositan* (ANTAO et al., 2008). Locos sob seleção ou que estão ligados à regiões sob seleção devem apresentar valores de F_{ST} que desviam significativamente (*outlying*) do esperado sob modelo neutro (BEAUMONT, 2005). Uma distribuição nula baseada em 100.000 locos simulados foi gerada com base nos valores médios de F_{ST} obtidos pela distribuição empírica com base nos 5.871 locos SNPs avaliados. Com base nos percentis de 0,025 e 0,975 foram construídos os limites do intervalo de confiança de 95% para a distribuição esperada da relação entre heterozigosidade e F_{ST} . Locos foram considerados como *outliers*

quando suas estimativas de F_{ST} plotadas com relação às heterozigosidades recaíram fora dos limites do intervalo de confiança. A distribuição foi iterativamente modificada pela sucessiva retirada dos locos com comportamento *outlier* até que não fossem mais identificados locos com comportamento *outlier*. Este processo foi realizado considerando os grupos de variedades de mandiocas mansas, bravas e silvestres, e separadamente para variedades mansas e bravas. É importante ressaltar que desvios significativos do comportamento neutro detectados nesta análise podem estar relacionados tanto à influência da seleção positiva (locos acima do intervalo de confiança), quanto de seleção balanceadora (locos abaixo do intervalo de confiança, BEAUMONT; NICHOLS, 1996). Enquanto a seleção positiva resulta no favorecimento de alelos vantajosos, colaborando para a divergência entre populações, a seleção balanceadora resulta na manutenção de polimorfismos em populações, colaborando para que as frequências alélicas em populações diferentes sejam similares (ALLENDORF; LUIKART; AITKEN, 2013; BEAUMONT; NICHOLS, 1996). Desta forma, deste ponto em diante serão chamados de SNPs *outliers* apenas aqueles que recaíram acima do limite superior do intervalo de confiança. Deste ponto em diante, o termo SNP neutro será utilizado tanto para os marcadores que recaíram dentro do intervalo de confiança como para aqueles que ficaram abaixo do limite inferior. As demais análises descritas abaixo foram realizadas para o conjunto total de marcadores SNPs, e separadamente para os marcadores SNPs com comportamento neutro, e ainda considerando apenas os marcadores SNPs com comportamento *outlier*. Este procedimento foi realizado para se avaliar possíveis distinções entre os resultados considerando todos os marcadores, apenas aqueles que apresentam comportamento neutro e apenas os que hipoteticamente estão sob influência de seleção positiva.

Devido a limitações de processamento, análises espaciais da distribuição da diversidade genômica realizadas com o programa DIVA-GIS (www.diva-gis.org) puderam levar em consideração apenas os locos *outliers* detectados entre variedades de mandioca mansas e bravas. Considerando as 69 variedades mansas avaliadas com 202 locos SNPs hipoteticamente sob seleção, foram analisados o total de 27.876 alelos georreferenciados (69 indivíduos x 202 locos x 2 alelos por loco). Além disso, foram analisados 28.684 alelos georreferenciados considerando as 71 variedades bravas avaliadas com 202 locos SNPs hipoteticamente sob seleção. Foram gerados *grids* com base no número de alelos diferentes presente em células com diâmetro de 10 minutos (correspondendo à aproximadamente 18 Km da área de estudo). Para considerar a diversidade alélica presente nas células adjacentes, foi usada a opção de *circular neighborhood* com diâmetro de 0,45045 graus (correspondendo à aproximadamente 50 Km da área de estudo).

O relacionamento dos indivíduos e dos grupos de variedades foi avaliado por meio da construção de dendrogramas da mesma forma como descrito para os marcadores ncSSR. Para avaliar a estrutura genética existente entre os grupos de variedades de mandioca foram realizadas análises multivariadas por meio de DAPCs realizadas com o pacote da plataforma R (R DEVELOPMENT CORE TEAM, 2003) *Adegenet* v.1.3.1 (JOMBART; AHMED, 2011). Para a representação geográfica dos agrupamentos identificados pelas DAPCs, a interpolação espacial dos coeficientes de adesão aos grupos foi realizada pelo método Kriging, para cada um dos agrupamentos identificados. Foi utilizada uma função disponibilizada por Oliver François (*Université Joseph Fourier, Institut National Polytechnique de Grenoble*) na plataforma R (R DEVELOPMENT CORE TEAM, 2003).

A partir das análises de estrutura genética realizadas com as DAPCs foram identificados os alelos e locos SNPs que mais contribuíram para a diferenciação entre os grupos encontrados. A posição genômica destes locos foi descoberta com base no arquivo VCF. Com o auxílio da ferramenta *on-line* *JBrowse* (SKINNER et al., 2009) inspeções manuais foram realizadas no genoma parcial de mandioca *Manihot esculenta* v4.1 (PROCHNIK et al., 2012) disponível no banco de dados *Phytozome* v.10.2 (<http://www.phytozome.net/cassava.php>), em busca de marcadores SNPs associados à regiões transcritas. Quando disponíveis, anotações funcionais depositadas nos bancos de dados do JGI (*Joint Genome Institute*) e instituições parceiras foram acessadas com a ferramenta *on-line* *PhytoMine* v.10.2.1, uma versão da ferramenta *InterMine* (SMITH et al., 2012; KALDERIMIS et al., 2014) customizada para usuários da plataforma *Phytozome*. Para as anotações funcionais foram considerados os termos GO (*Gene Ontology*), que resumizam a informação sobre os componentes celulares, funções moleculares e processos biológicos nos quais os produtos funcionais dos transcritos identificados estão envolvidos (BLAKE, 2013).

4 RESULTADOS

4.1 Diversidade e estrutura haplotípica de variedades de mandioca com base em cpSSR

Dos 20 locos cpSSR testados, 16 apresentaram padrões de amplificação claros, mas apenas quatro foram polimórficos, revelando dois ou três alelos (Tabela 2). As variedades de mandiocas bravas apresentaram índices de diversidade similares às mandiocas mansas. Ambos os grupos de variedades cultivadas apresentaram índices de diversidade maiores que os indivíduos silvestres (Tabela 3). Os grupos de indivíduos de diferentes rios apresentaram índices de diversidade semelhantes, com exceção das variedades cultivadas ao longo do rio Solimões que apresentaram estimativas do índice de Shannon e de diversidade gênica inferior aos demais rios e similares às variedades do rio Guaporé e do nordeste do Pará (Tabela 3).

Tabela 2 - Estimativas de diversidade genética para os locos cpSSR polimórficos em 596 indivíduos de mandioca. A = número de alelos, I = Índice de Shannon, h = diversidade gênica

Loco	A	I	h	Referência
ccmp05	3	0,561	0,289	Weising & Gardner (1999)
ccmp06	3	0,567	0,360	Weising & Gardner (1999)
ccmp10	2	0,023	0,007	Weising & Gardner (1999)
ccSSR7	3	0,320	0,166	Chung & Staub (2003)
Média	2,75	0,367	0,205	

Tabela 3 - Estimativas de diversidade genética para os locos cpSSR, considerando os grupos de variedades cultivadas e os indivíduos silvestres, e os grupos de variedades cultivadas ao longo dos rios e regiões. $\bar{A}$ = número médio de alelos, I = Índice de Shannon, h = diversidade gênica

Grupos de variedades	$\bar{A}$	I	h
Silvestres	1,5	0,077	0,034
Bravas	2,25	0,306	0,170
Mansas	2,75	0,291	0,147
Rios / Regiões			
Madeira	2	0,283	0,153
Negro/Branco	2,25	0,341	0,193
Solimões	1,5	0,124	0,063
Amazonas	2,75	0,291	0,136
Guaporé	1,75	0,139	0,065
Pará	1,5	0,122	0,063

Quando avaliados em conjunto, o total de quatro locos cpSSR polimórficos resultaram em 11 haplótipos cloroplastidiais diferentes (Tabela 4), dos quais três ocorreram em apenas um indivíduo, sendo que dois destes haplótipos foram encontrados apenas em variedades mansas, e um foi encontrado apenas em uma variedade brava.

Tabela 4 - Haplótipos cloroplastidiais obtidos para variedades cultivadas e silvestres, com base em quatro locos cpSSR

Haplótipo	ccmp05	ccmp06	ccmp10	ccSSR7	contagem
A	136	109	123	363	7
B	136	109	123	377	45
C	137	109	123	363	48
D	137	110	123	363	430
E	137	110	123	377	5
F	138	108	125	368	2
G	138	109	123	363	36
H	138	110	123	363	5
1	136	108	123	377	1
2	136	110	123	363	1
3	136	110	123	377	1

Apesar do menor número de indivíduos as variedades mansas apresentaram o mesmo número de haplótipos que as variedades bravas. Entretanto, as demais estimativas de diversidade haplotípica foram um pouco mais elevadas para as variedades bravas. A diversidade haplotípica foi maior nos grupos de variedades cultivadas em comparação aos indivíduos silvestres (Tabela 5). Os grupos de variedades cultivadas ao longo dos rios apresentaram estimativas de diversidade haplotípica similares, com exceção das variedades cultivadas ao longo do rio Solimões, que apresentaram índices mais baixos e comparáveis ao conjunto de variedades provenientes do rio Guaporé e do nordeste do Pará (Tabela 5).

A avaliação da rede de haplótipos demonstrou a inexistência de associação entre grupos de haplótipos e variedades de mandioca mansa, brava ou silvestre. Dos 11 haplótipos encontrados (desconsiderando os indivíduos não designados) três foram encontrados apenas em variedades bravas, outros três apenas em variedades mansas, e os demais foram compartilhados entre variedades mansas e bravas, ou entre estas e indivíduos silvestres (Figura 11A). Também não foram encontradas associações entre grupos de haplótipos e as regiões fluviais em que as variedades de mandioca foram coletadas (Figura 11B). Notavelmente o haplótipo B, mais comum entre indivíduos silvestres, foi compartilhado com variedades mansas e bravas provenientes dos rios Amazonas, Negro/Branco, Madeira e ainda do estado do Pará.

Tabela 5 - Estimativas de diversidade haplotípica para variedades de mandioca cultivada e indivíduos silvestres, e grupos de variedades cultivadas ao longo dos rios. N = número de indivíduos, N_H = número de haplótipos, Ne_H = número efetivo de haplótipos, *I* = Índice de Shannon, H_{Ehap} = diversidade haplotípica, π = diversidade nucleotídica

Grupo de variedades	N	N_H	Ne_H	<i>I</i>	H_{Ehap}	π
Bravas	318	8	1,72	0,942	0,420	0,124
Mansas	218	8	1,46	0,733	0,318	0,107
Cultivadas ‡	553	11	1,62	0,889	0,383	0,119
Silvestres	28	2	1,07	0,154	0,071	0,026
Rios / Regiões						
Madeira	164	7	1,49	0,779	0,335	0,112
Negro/Branco	206	8	1,79	0,949	0,443	0,141
Solimões	74	2	1,14	0,247	0,128	0,046
Amazonas	65	6	1,29	0,557	0,231	0,100
Guaporé	22	3	1,20	0,368	0,177	0,049
Pará	22	3	1,20	0,368	0,177	0,048
Rios (variedades) / Regiões (variedades)						
Madeira bravas	108	7	1,62	0,894	0,388	0,121
Madeira mansas	49	3	1,23	0,381	0,190	0,076
Negro/Branco bravas	117	7	2,11	1,073	0,530	0,161
Negro/Branco mansas	86	6	1,41	0,678	0,295	0,107
Solimões bravas	56	2	1,19	0,301	0,166	0,060
Solimões mansas	18	1	1,00	0,000	0,000	0,000
Amazonas bravas	26	2	1,08	0,163	0,077	0,014
Amazonas mansas	39	5	1,46	0,700	0,325	0,153
Guaporé bravas	2	1	1,00	0,000	0,000	0,000
Guaporé mansas	13	1	1,00	0,000	0,000	0,000
Pará bravas	9	2	1,25	0,349	0,222	0,056
Pará mansas	13	2	1,17	0,271	0,154	0,056

‡ Considerando variedades mansas, bravas e não-designadas

Figura 11 - Rede de haplótipos observados em variedades cultivadas e silvestres de mandioca. Cada círculo representa um haplótipo, e o tamanho dos círculos é proporcional à frequência do haplótipo. Diferentes cores representam em A) mandiocas mansas, bravas e silvestres, e em B) mandiocas por rio ou região de coleta. Cada barra transversal representa uma mutação entre os haplótipos

A distribuição espacial dos haplótipos pode ser visualizada na Figura 12. De modo geral, considerando todos os indivíduos pode-se perceber a ampla distribuição do haplótipo D ao longo dos principais rios da bacia Amazônica, contribuindo para um padrão mais ou menos homogêneo entre os rios e regiões (Figura 12A). Exceção clara são os indivíduos provenientes do Pará, e do município de São Gabriel da Cachoeira (AM), no alto rio Negro, que apresentam em quase sua totalidade indivíduos com haplótipos C, que diferem do haplótipo D por apenas uma mutação. Além de seguir as tendências gerais, a distribuição espacial dos haplótipos das variedades de mandioca bravas indica uma área de maior diversidade haplotípica na região ao redor do rio Branco, no estado de Roraima, onde os haplótipos G e B (compartilhado com indivíduos silvestres) foram frequentemente encontrados (Figura 12B). As tendências gerais foram também observadas para as variedades de mandioca mansa (Figura 12C).

Figura 12 - Distribuição geográfica dos haplótipos observados em variedades cultivadas e silvestres de mandioca amostrados em diferentes municípios (pontos pretos) ao longo dos principais rios da bacia Amazônica brasileira, e no leste do estado do Pará. Cada haplótipo é identificado por uma cor diferente indicada na rede de haplótipos na caixa. Distribuição dos haplótipos considerando A) todos os indivíduos, B) mandiocas bravas e C) mandiocas mansas. O tamanho dos círculos não está relacionado ao número de amostras em cada local, mas alguns estão aumentados para melhor visualização

A análise da distribuição da variabilidade genética em grupos feita por meio da AMOVA revelou que a maior parte da variação quando consideradas as variedades mansas e bravas está dentro dos grupos de variedades, já que a divergência entre estes grupos foi muito baixa, com $F_{ST} = 0,005$ (Tabela 6). Quando os indivíduos silvestres são comparados com as variedades cultivadas, a maior parte da variação passa a ser observada entre os grupos, com $F_{ST} = 0,78$. A divergência entre variedades cultivadas agrupadas de acordo com os rios foi baixa, mas

significativa, e mais alta do que a comparação entre variedades mansas e bravas, com $F_{ST} = 0,028$. Porém, quando a variação presente entre rios é desmembrada, pode-se observar que há uma grande contribuição dos grupos de variedades mansas e bravas ($F_{CT} = 0,013$). Quando as variedades provenientes do nordeste do Pará são adicionadas à análise por regiões, a estimativa de diferenciação aumenta ($F_{CT} = 0,078$), sugerindo que as variedades provenientes desta região apresentam alto grau de divergência com relação às variedades cultivadas ao longo dos principais rios amazônicos, quase certamente devido à predominância de haplótipos C.

Tabela 6 - Análise de variância molecular (AMOVA) baseada em quatro locos cpSSR para diferentes grupos de variedades de mandioca. G.L. = graus de liberdade. Valores em negrito indicam as estatísticas correspondentes a diferenciação entre grupos (F_{ST}) em cada AMOVA

Fonte de variação	G.L.	Porcentagem da variação (%)	Estatísticas F
Entre mansas e bravas	1	0,50	$F_{ST} = 0,005$
Dentro de mansas e bravas	534	99,50	
Entre cultivadas e silvestres	1	78,00	$F_{ST} = 0,780^*$
Dentro de cultivadas e silvestres	579	22,00	
Entre rios	4	2,87	$F_{ST} = 0,028^*$
Dentro de rios	526	97,13	
Entre rios [†]	4	1,36	$F_{CT} = 0,013^*$
Entre variedades dentro de rios	8	2,89	$F_{SC} = 0,029$
Dentro de variedades	518	95,76	$F_{ST} = 0,042^*$
Entre regiões [§]	5	7,83	$F_{CT} = 0,078^*$
Entre municípios dentro de regiões	38	20,22	$F_{SC} = 0,219^*$
Dentro de municípios	512	71,95	$F_{ST} = 0,280^*$

*Significativo com $p < 0,05$.

[†] Considerando mandiocas provenientes dos rios Madeira, Negro/Branco, Solimões, Amazonas e Guaporé.

[§] Considerando mandiocas provenientes dos rios e do nordeste do estado do Pará.

O número de agrupamentos mais provável encontrado a partir da análise Bayesiana realizada no programa *Structure* foi $K = 3$, seguido por $K = 4$ (Figura 13). Entretanto, o que se observa é um padrão com dois agrupamentos: o primeiro com os indivíduos silvestres e variedades mansas e bravas provenientes da maioria das localidades de coleta, e o segundo com variedades mansas e bravas provenientes dos 43 municípios. Apesar disso, não houve correspondência entre os agrupamentos formados pela análise com os grupos de variedades mansas e bravas, nem com relação aos principais rios. Interessantemente, quando considerado $K = 3$ as variedades de São Gabriel da Cachoeira e do oeste do estado do Pará, que apresentaram predominância do haplótipo C, parecem ser mais relacionadas aos indivíduos silvestres do que a maioria das variedades mansas e bravas dos outros municípios (Figura 13B). Entretanto, este relacionamento mais próximo se torna menos evidente quando se considera $K = 4$ (Figura 13C).

Figura 13 - Estrutura genética sugerida pelo programa *Structure* para variedades de mandioca, com base em quatro locos cpSSR. A) Gráfico da estatística ΔK obtidos a partir de 10 simulações independentes para cada número K de agrupamentos. B) Gráficos de barras representando as probabilidades de ancestralidade de 596 indivíduos de mandioca considerando $K = 3$ e C) $K = 4$. Cada barra representa um indivíduo, e cores diferentes representam as probabilidades dos indivíduos de pertencerem a grupos genéticos distintos. As linhas pretas separam os indivíduos de acordo com os municípios. Asteriscos indicam os indivíduos silvestres

Pelo método de *K-means* foram encontrados três agrupamentos de indivíduos de mandioca e por meio da DAPC o relacionamento entre agrupamentos e indivíduos pode ser avaliada (Tabela 7). Os dois primeiros componentes principais retiveram 99,38% da variação haplotípica total, e os resultados foram similares aos obtidos pela análise Bayesiana realizada no programa *Structure*.

Um dos agrupamentos foi formado pela maioria dos indivíduos silvestres, com alguns indivíduos de variedades bravas e mansas (Tabela 7 e Figura 14A). Os outros dois agrupamentos consistiram de mistura de indivíduos de ambos os conjuntos de variedades cultivadas. Pode-se observar ainda, a distinção dos indivíduos silvestres, com alguns indivíduos de variedades cultivadas mais próximos a estes, e uma grande sobreposição dos grupos de variedades mansas e bravas (Tabela 7 e Figura 14B). Não houve correspondência dos agrupamentos formados com os rios e regiões de origem.

Tabela 7 - Distribuição dos indivíduos de mandioca entre os agrupamentos formados pela Análise Discriminante de Componentes Principais (DAPC), baseado em quatro locos cpSSR, considerando os grupos de variedades, rios e regiões de coleta

Variedades	Agrupamentos DAPC		
	1	2	3
Bravas	254	11	60
Mansas	189	8	29
ND	13	1	3
Silvestres	-	27	1
Rios/Regiões			
Madeira	145	4	20
Negro/Branco	157	14	39
Solimões	71	-	5
Amazonas	62	1	6
Guaporé	21	-	1
Pará	-	1	21
Silvestres	-	27	1

Figura 14 - Análise Discriminante de Componentes Principais (DAPC) com base em quatro locos cpSSR para 596 indivíduos de mandioca. A) Gráfico de dispersão dos agrupamentos formados pelo método de *K-means*; B) Gráfico de densidade dos indivíduos de acordo com grupos de variedades cultivadas e silvestres ao longo da primeira função discriminante obtida na DAPC

4.2 Diversidade e estrutura genética de variedades de mandioca com base em ncSSR

Todos os 15 locos ncSSR testados foram amplificados com sucesso, mas o loco GA127 não apresentou perfis claros para a determinação dos alelos e foi desconsiderado. Os demais 14 locos puderam ser arranjados em sistemas *multiplex* para a detecção do polimorfismo, de acordo com a modificação fluorescente disponível e com o tamanho esperado dos fragmentos amplificados (Tabela 8). O número médio de alelos encontrado por loco foi de 9, variando de

quatro (GA12, SSRY32 e 89) a 18 (SSRY 164). As heterozigosidades médias observada e esperada foram próximas (0,616 e 0,650), sendo que cinco dos 14 locos apresentaram heterozigosidades observadas (H_O) maiores que as esperadas (H_E), resultando em coeficientes de endogamia (f) negativos. Três locos apresentaram valores de H_O e H_E aproximados, e outros seis apresentaram valores de H_O inferiores aos de H_E (Tabela 8).

Tabela 8 - Características e estimativas de diversidade genética para 14 locos ncSSR avaliados em 596 indivíduos de mandioca. Características: organização em sistemas *multiplex*, modificação fluorescente, e variação do tamanho de fragmentos amplificados em pares de base (pb). Estimativas de diversidade genética: número de alelos (A), heterozigosidades observada (H_O) e esperada (H_E), e estimativa do coeficiente de endogamia (f). *Significativo com intervalo de confiança de 95%

<i>multiplex</i>	loco	fluorescência	tamanho (pb)	A	H_O	H_E	f
1	GA136	6-FAM	140-158	8	0,705	0,719	0,019
	GA21	HEX	104-118	6	0,419	0,571	0,266*
	GA134	HEX	308-332	8	0,374	0,470	0,205*
	GA126	NED	176-216	10	0,776	0,764	-0,016
2	GA140	6-FAM	148-172	10	0,632	0,747	0,154*
	GA131	HEX	88-116	12	0,711	0,801	0,113*
	GAGG5	NED	111-125	6	0,650	0,555	-0,172*
	SSRY32	NED	299-305	4	0,648	0,583	-0,112*
3	SSRY164	6-FAM	149-191	18	0,700	0,792	0,117*
	SSRY89	HEX	104-118	4	0,128	0,133	0,042
	SSRY13	NED	196-238	12	0,706	0,845	0,165*
4	GA12	6-FAM	127-143	4	0,645	0,571	-0,131*
	Y20	NED	124-158	13	0,770	0,765	-0,006
	Y70	NED	222-254	11	0,759	0,777	0,024
Média				9	0,616	0,650	0,048*

Os 596 indivíduos de mandioca avaliados com 14 locos ncSSR apresentaram um total de 395 genótipos multilocos (MLGs, que correspondem à combinações únicas de alelos). Todos os indivíduos silvestres apresentaram MLGs diferentes. Além disso, o número de MLGs foi proporcionalmente maior nas variedades bravas do que nas variedades mansas. O número médio de alelos e a riqueza alélica foi um pouco maior para as variedades de mandioca brava, porém as variedades mansas apresentaram maior proporção de heterozigotos (Tabela 9). Com isso as variedades de mandioca mansa apresentaram significativo excesso de heterozigotos enquanto as variedades bravas apresentaram um pequeno, mas também significativo déficit ($f = -0,157$ e $0,065$, respectivamente). As variedades cultivadas apresentaram número médio de alelos maior do que os indivíduos silvestres, além de maiores heterozigosidades. Entretanto, quando se leva em consideração a riqueza alélica, os indivíduos silvestres apresentaram diversidade alélica similar às variedades cultivadas. Os indivíduos silvestres apresentaram também maior número de alelos privados do que as variedades de mandioca mansas e bravas.

Em contrapartida, os indivíduos silvestres apresentaram o maior déficit de heterozigotos entre todos os grupos.

Tabela 9 - Estimativas de diversidade genética para 14 locos ncSSR avaliados em 596 indivíduos de mandioca, considerando variedades de mandioca cultivada e indivíduos silvestres, variedades cultivadas ao longo dos rios ou regiões, e grupos de variedades mansas e bravas ao longo dos rios ou regiões. Número de indivíduos (N), número de genótipos multilocos (MLGs), número médio de alelos ($\bar{A}$), riqueza alélica (A_R), número de alelos privados (Ap), heterozigosidades observada (H_O) e esperada (H_E), e estimativa do coeficiente de endogamia (f). *Significativo com intervalo de confiança de 95%

Grupos de variedades	N	MLGs	$\bar{A}$	A_R	Ap	H_O	H_E	f
Bravas	325	247	6,79	4,75	9	0,573	0,614	0,065*
Mansas	226	121	6,64	4,33	10	0,691	0,595	-0,157*
Cultivadas ‡	568	376	8,28	8,28	58	0,626	0,635	0,009
Silvestres	28	28	4,86	4,28	13	0,466	0,521	0,122
Rios / Regiões								
Madeira	169	127	6,36	5,07	11	0,638	0,636	0,004
Negro/Branco	210	129	6,36	4,89	5	0,602	0,616	0,018
Solimões	76	59	5,50	4,88	-	0,587	0,611	0,021
Amazonas	69	49	6,29	5,25	6	0,640	0,624	0,008
Guaporé	22	11	4,57	4,54	3	0,763	0,598	-0,263*
Pará	22	13	4,21	4,19	-	0,638	0,613	-0,053
Variedades por rio / regiões								
Madeira bravas	111	90	5,36	2,43	3	0,619	0,629	0,013
Madeira mansas	51	36	5,00	2,29	4	0,701	0,588	-0,163*
Negro/Branco bravas	119	86	5,50	2,24	2	0,540	0,566	0,046
Negro/Branco mansas	88	45	5,21	2,25	1	0,681	0,580	-0,167*
Solimões bravas	58	46	5,14	2,24	-	0,535	0,563	0,035
Solimões mansas	18	13	3,64	2,25	-	0,754	0,578	-0,303*
Amazonas bravas	26	23	5,36	2,39	2	0,594	0,610	0,074
Amazonas mansas	43	26	5,00	2,25	4	0,667	0,571	-0,159*
Guaporé bravas	2	1	1,64	1,64	-	0,643	0,321	-1,000*
Guaporé mansas	13	8	3,93	2,34	2	0,753	0,613	-0,230*
Pará bravas	9	4	3,29	2,21	-	0,611	0,567	-0,066
Pará mansas	13	10	3,93	2,35	-	0,657	0,616	-0,082

‡ Considerando variedades mansas, bravas e não-designadas

Quando as variedades são avaliadas de acordo com os rios ou regiões ao longo dos quais foram coletadas, estimativas de diversidade um pouco menores são observadas para o rio Solimões em comparação aos outros rios mais importantes (Madeira, Negro/Branco e Amazonas). Pode ser observado um número elevado de MLGs com relação ao total de indivíduos amostrados para as variedades bravas dos principais rios (Madeira, Negro/Branco, Solimões e Amazonas). Este quadro se inverte para as outras regiões (rio Guaporé e nordeste do Pará), onde são as variedades mansas que apresentam números elevados de MLGs em relação ao total de indivíduos amostrados nestas regiões (Tabela 9). As variedades bravas dos principais rios apresentam ligeiro déficit de heterozigotos, enquanto que as variedades mansas

apresentam marcantes excessos de heterozigotos (Tabela 9). Apesar do número médio de alelos ser menor do que os principais rios, as variedades provenientes do rio Guaporé e do nordeste do Pará apresentaram elevadas proporções de heterozigotos ($H_o = 0,763$ e $0,638$, respectivamente), sendo os únicos agrupamentos por rios ou regiões que apresentaram excesso de heterozigotos (f negativo, Tabela 9).

A Figura 15 apresenta a distribuição geográfica dos alelos encontrados com os 14 locos ncSSR avaliados. Considerando todos os indivíduos amostrados, apenas duas pequenas áreas localizadas no oeste de Rondônia (alto rio Madeira) e no centro de Roraima (médio rio Branco) concentram um número total de alelos mais elevado em relação às demais localidades (Figura 15A, esquerda). Além disso, maior número de alelos pode ser observado ao longo dos rios Madeira e Negro/Branco em relação aos rios Solimões e Amazonas. Porém, quando é feita a rarefação do número de alelos (riqueza alélica) observa-se que os rios possuem níveis de diversidade genética similares (Figura 15A, direita). As variedades de mandioca bravas apresentaram maior número total de alelos ao longo dos rios Madeira e Negro (Figura 15B, esquerda). Entretanto, através da distribuição das riquezas alélicas das variedades bravas percebe-se que a maior parte das localidades situadas ao longo dos principais rios apresentam níveis intermediários de diversidade genética (Figura 15B, direita). Considerando as variedades de mandioca mansa, regiões com elevado número total de alelos estão presentes no oeste de Rondônia (alto rio Madeira) e ao longo do rio Branco em Roraima (Figura 15C, esquerda). A distribuição das riquezas alélicas das variedades mansas indica que a maior parte das localidades situadas aos longos dos principais rios apresentam níveis intermediários de diversidade genética (Figura 15C, direita), estando as localidades com riquezas alélicas mais elevadas ao longo do alto e médio rio Negro.

Figura 15 - Mapas mostrando o número total de alelos [coluna da esquerda, em células de 10 minutos (aproximadamente 18 Km) aplicando o método de *circular-neighborhood* de 0,45045 graus (aproximadamente 50 Km)], e a riqueza alélica (coluna da direita, em células de 0,45045 graus), com base na variação genética revelada por 14 marcadores ncSSR. Em A) todos os indivíduos, em B) apenas variedades de mandioca brava, e em C) apenas variedades de mandioca mansa

O número de agrupamentos mais provável encontrado a partir da análise Bayesiana realizada no programa *Structure* com todos os 596 indivíduos de mandioca foi $K = 2$, nos quais os indivíduos silvestres apresentam probabilidades semelhantes de pertencerem aos dois grupos sugerido pela análise (Figura 16A). Houve uma concordância razoável entre os agrupamentos e as variedades mansas e bravas. Considerando o limite arbitrário de pelo menos 80% de probabilidade de ancestralidade em um dos agrupamentos ($q \geq 0,8$), 267 (82%) das 325 mandiocas bravas foram alocadas no grupo 1, enquanto 143 (63%) das 226 variedades mansas

foram alocados no grupo 2. Quando a matriz q de ancestralidade é interpolada espacialmente em mapas, observa-se uma estrutura complexa (Figura 16A) para os grupos de variedades cultivadas. O primeiro grupo é formado por variedades provenientes do alto e médio rio Negro (noroeste do estado do Amazonas), que são mais relacionadas às variedades do médio e baixo rio Solimões e médio e baixo rio Madeira. O segundo grupo é formado principalmente por variedades provenientes do leste do estado do Pará, sul do estado de Rondônia, extremo oeste do estado do Amazonas (município de Parintins, no rio Amazonas) e sul de Roraima (rio Branco).

Para a análise do programa *Structure* com as 325 mandiocas bravas, o número de agrupamentos mais provável também foi de $K = 2$ (Figura 16B). Considerando o limite arbitrário de pelo menos 80% de probabilidade de ancestralidade em um dos agrupamentos, 83 (25,5%) indivíduos foram alocados no primeiro grupo e 163 (50,1%) foram alocados no outro. Na interpolação espacial observa-se que a maioria das localidades amostradas apresentam níveis intermediários de ancestralidade ($q < 0,8$). O primeiro grupo é formado por variedades do leste do estado do Pará, sul (região do rio Guaporé) e oeste (alto rio Madeira) de Rondônia, além do norte de Roraima (rio Branco). No segundo grupo destacam-se as variedades provenientes do alto e médio rio Negro (noroeste do estado do Amazonas) e do alto rio Amazonas (porção centro-leste do estado do Amazonas).

Considerando apenas as 226 mandiocas mansas, o número de agrupamentos mais provável detectado pelo programa *Structure* foi de $K = 3$ (Figura 16C). Levando em conta o limite arbitrário de pelo menos 80% de probabilidade de ancestralidade em um dos agrupamentos, o número de indivíduos em cada um dos três grupos foi similar, sendo de aproximadamente 20%, 25% e 26%. Os demais 29% dos indivíduos podem ser considerados como mistura dos três grupos. Isto resulta em uma distribuição espacial dos coeficientes de ancestralidade mais homogênea para as variedades mansas, com uma pequena tendência de maior relacionamento entre as variedades provenientes dos rios Solimões e Amazonas (Figura 16C).

Figura 16 - Gráficos de barra e interpolação espacial dos coeficientes de ancestralidade (q) para os agrupamentos identificados pela análise de estruturação genética obtida com o programa *Structure*. Mapas são mostrados para cada um dos K agrupamentos. A barra de escala à direita dos mapas representa a probabilidade dos indivíduos (pontos pretos) de pertencerem ao agrupamento. Estão representadas as análises com: A) todos os 596 indivíduos (asterisco indica os indivíduos silvestres), B) com 325 indivíduos de mandiocas bravas, e C) com 226 indivíduos de mandiocas mansas

As DAPCs revelaram essencialmente os mesmos resultados que a análise Bayesiana. Por meio do método de *K-means* o número de agrupamentos que melhor explica os dados de variação alélica dos 14 locos ncSSR foi $K = 7$. Entretanto a DAPC também foi realizada para três grupos de variedades (mansas, bravas e silvestres), por estes representarem com acurácia e

de forma mais simples a divisão dos indivíduos amostrados (Tabela 10).

Tabela 10 - Distribuição dos indivíduos de mandioca entre os agrupamentos formados pela Análise Discriminante de Componentes Principais (DAPC), baseado em 14 locos ncSSR, considerando os grupos de variedades, rios e regiões de coleta

Variedades	DAPC 3 agrupamentos			DAPC 7 agrupamentos						
	1	2	3	1	2	3	4	5	6	7
Silvestres	27	-	1	26	-	-	1	1	-	-
Bravas	-	34	291	-	7	123	16	72	68	39
Mansas	-	162	64	-	75	11	46	10	1	83
ND	-	10	7	-	-	2	-	-	1	14
Rios / Regiões										
Madeira	-	70	99	-	22	28	21	14	30	54
Negro/Branco	-	52	158	-	38	50	8	51	29	34
Amazonas	-	33	36	-	17	9	15	8	9	11
Solimões	-	19	57	-	3	45	9	8	2	9
Guaporé	-	19	3	-	2	-	3	-	-	17
Pará	-	13	9	-	-	4	6	1	-	11
Silvestres	27		1	26	-	-	1	1	-	-
Total Geral	27	206	363	26	82	136	63	83	70	136

A DAPC que considerou 7 agrupamentos *a priori* resultou em um grupo de indivíduos silvestres bem definido, e com uma subdivisão dos grupos de variedades cultivadas nos demais seis grupos (Figura 17A). Para a DAPC considerando três agrupamentos *a priori*, os indivíduos silvestres formaram um grupo bem distinto (Figura 17B), e os outros dois grupos tiveram melhor correspondência com mandiocas bravas e mansas do que os resultados do programa *Structure*. Dentre as mandiocas mansas 71,7% foram alocados no grupo 2, enquanto 89,5% das variedades bravas foram alocados no grupo 3. Quando apenas a primeira função discriminante da análise com sete agrupamentos é retida e representada de acordo com os grupos de variedades (Figura 17C), observa-se a sobreposição de variedades de mandioca mansas e bravas.

As AMOVAs revelaram que a maior parte da variação ocorre dentro dos grupos de indivíduos avaliados. As variedades mansas e bravas apresentaram moderada divergência genética, com $F_{ST} = 0,083$ (Tabela 11). Quando os indivíduos silvestres são comparados com as variedades cultivadas o valor de F_{ST} (0,28) foi mais baixo do que o encontrado para cpSSRs (0,78), entretanto este valor ainda é significativamente alto. A diferença entre os grupos de variedades cultivadas de acordo com os rios foi baixa, mas significativa, com $F_{ST} = 0,025$. Por outro lado, há um aumento na divergência entre grupos ($F_{CT} = 0,079$) quando se considera a variação presente entre variedades mansas e bravas agrupadas por rio (Tabela 11).

Figura 17 - Análise Discriminante de Componentes Principais (DAPC) com base em 14 locos ncSSR para 596 indivíduos de mandioca. A) Gráfico de dispersão dos sete agrupamentos formados pelo método de *K-means*; B) Gráfico de dispersão considerando três agrupamentos; C) Gráfico de densidade dos indivíduos de acordo com grupos de variedades cultivadas e silvestres ao longo da primeira função discriminante obtida na DAPC

Tabela 11 - Análises de variância molecular (AMOVA) baseada na média de 14 locos ncSSR para diferentes grupos de variedades de mandioca. Valores em negrito indicam as estatísticas correspondentes a diferenciação entre grupos (F_{ST}) em cada AMOVA

Fonte de variação	Soma de Quadrados	Coefficiente de Variação	Porcentagem da Variação (%)	Estatísticas F
Entre mansas e bravas	209,35	0,39	8,35	$F_{ST} = \mathbf{0,083^*}$
Dentro de mansas e bravas	4.643,85	4,24	91,65	
Entre cultivadas e silvestre	181,19	1,70	27,89	$F_{ST} = \mathbf{0,279^*}$
Dentro de cultivadas e silvestre	5.206,23	4,40	72,11	
Entre rios	105,13	0,11	2,51	$F_{ST} = \mathbf{0,025^*}$
Dentro de rios	4.708,34	4,35	97,49	
Entre rios ‡	213,63	0,37	7,88	$F_{CT} = \mathbf{0,079^*}$
Entre variedades mansas e bravas dentro de rios	154,56	0,16	3,53	$F_{SC} = 0,038^*$
Dentro de variedades	4.288,69	4,11	88,59	$F_{ST} = 0,114^*$
Entre regiões §	118,71	0,07	1,58	$F_{CT} = \mathbf{0,016^*}$
Entre municípios dentro de regiões	351,90	0,21	4,78	$F_{SC} = 0,048^*$
Dentro de municípios	4.581,97	4,20	93,64	$F_{ST} = 0,063^*$

*Significativo com $p < 0,05$.

‡ Considerando mandiocas provenientes dos rios Madeira, Negro/Branco, Solimões, Amazonas e Guaporé.

§ Considerando mandiocas provenientes dos rios e do nordeste do estado do Pará.

As elevadas divergências entre variedades cultivadas e os indivíduos silvestres, além dos diferentes graus de divergência entre variedades provenientes das calhas dos rios podem ser também verificadas por meio da análise dos valores de F_{ST} par-a-par (Figura 18). A maioria dos valores de F_{ST} entre as variedades mansas e bravas provenientes dos diferentes rios e regiões foi baixa. De modo geral, os indivíduos provenientes das regiões do rio Guaporé e do nordeste do estado do Pará apresentaram os maiores valores de diferenciação genética (Figura 18B e C).

Figura 18 - Estimativas de F_{ST} (diagonal inferior) e *heatmap* dos valores de F_{ST} (diagonal superior), com base em 14 locos ncSSR, entre pares de: A) variedades mansas, bravas e silvestres; B) entre grupos de variedades de acordo com os rios ou regiões; C) entre variedades mansas (M) e bravas (B) de acordo com rios ou regiões. Valores em negrito são significativos com intervalo de confiança de 95%. As cores variam do branco (menores valores de F_{ST}) ao azul escuro (maiores valores de F_{ST})

A análise dos relacionamentos entre os 596 indivíduos de mandioca com base em um dendrograma mostrou agrupamentos semelhantes com os encontrados nas análises Bayesianas e multivariadas (Figura 19). As mandiocas silvestres encontram-se na base do dendrograma, formando um grupo consistente, no qual está inserido um indivíduo não designado do município de Humaitá (sudeste do estado do Amazonas). A seguir encontram-se dois grupos, sendo o primeiro formado por uma variedade mansa do município de Humaitá e uma variedade brava do município de Monte Alegre (estado do Pará). O segundo grupo é formado por variedades bravas dos municípios de Humaitá e Santo Antônio do Içá (oeste do estado do Amazonas). Logo após são observados três grandes agrupamentos: o primeiro constituído majoritariamente por variedades mansas, seguido por um grupo que reúne variedades mansas e bravas, e um terceiro

agrupamento mais interno no qual predominam variedades bravas. Não houve a formação de grupos consistentes quando os indivíduos são classificados de acordo com os rios e regiões de coleta.

Figura 19 - Dendrograma mostrando os relacionamentos entre 596 indivíduos de variedades de mandiocas construído pelo método *Neighbor-Joining*, com base em distâncias de Cavalli-Sforza & Edwards (1967) a partir de 14 locos ncSSR

A avaliação dos relacionamentos entre os grupos de variedades mansas e bravas de acordo com rios e regiões com base em um dendrograma mostrou que as variedades bravas da região do rio Guaporé foram as mais proximamente relacionadas aos indivíduos silvestres (Figura 20A). As variedades provenientes desta região são as que estão geograficamente mais próximas aos indivíduos silvestres. As mandiocas mansas provenientes dos principais rios amazônicos formaram um grupo consistente, bem como as variedades bravas, cujo grupo ficou na porção mais interna do dendrograma. Quando as variedades da região do rio Guaporé e do nordeste do estado do Pará são retiradas da análise as variedades mansas e bravas formam agrupamentos consistentes e com alto suporte (Figura 20B). Entretanto as mandiocas bravas deixam de ocupar a porção mais interna do dendrograma, estando igualmente relacionadas às

mandiocas mansas e aos indivíduos silvestres.

Figura 20 - Dendrogramas mostrando os relacionamentos entre grupos de mandiocas mansas e bravas de acordo com rios e regiões de coleta construído pelo método *Neighbor-Joining*, com base em distâncias de Cavalli-Sforza & Edwards (1967) a partir de 14 locos ncSSR. A) variedades agrupadas por rios e regiões de coleta. B) Variedades agrupadas por rios. Valores de *bootstrap* acima de 70% são apresentados

4.3 Diversidade e estrutura genômica de variedades de mandioca com base em marcadores SNPs

A aplicação da técnica RAD-seq em 159 variedades de mandioca resultou em um total de 412.504.046 sequências, com média de 2.562.136,9 sequências obtidas por indivíduo. Cada sequência teve uma profundidade média ao redor de 15X. Em média, 57,8% das sequências puderam ser alinhadas com o genoma parcial da mandioca, enquanto 39,7% não alinharam com nenhuma região e os demais 2,6% alinharam com mais de uma região ao longo do genoma. Foram obtidos 5.871 marcadores SNPs, com cobertura média de 535X, *Phred-score* médio de 1.104,8. A partir destes marcadores foram obtidos um total de 935.238 genótipos para os 159 indivíduos, dos quais 104.962 genótipos foram heterozigotos (11,22%), enquanto apenas 0,8% consistiam de dados faltantes.

A maioria dos marcadores SNPs apresentou heterozigosidades observadas comparáveis ou mesmo maiores que as heterozigosidades esperadas (Figura 21). Isto se reflete no coeficiente de endogamia (f) próximo a zero encontrado para a maioria dos marcadores SNPs (Figura 22). As mandiocas mansas e bravas apresentaram diversidade genômica similar, e um pouco maior do que os indivíduos silvestres (Tabela 12), mas as mandiocas mansas apresentaram maior número de alelos, e excesso de heterozigotos ($f = -0,015$) significativo. A diversidade genômica foi similar para os grupos de variedades por rio (ao redor de 11%). Com exceção das variedades

do rio Solimões, as mandiocas mansas apresentaram coeficientes de endogamia um pouco menores do que as mandiocas bravas provenientes dos demais rios.

A) silvestres

B) bravas

C) mansas

Figura 21 - Gráfico de dispersão ilustrando as médias das estimativas de heterozigosidades esperadas (H_E , azul) e observadas (H_O , preto) para os 5.871 marcadores SNPs detectados em 159 indivíduos de mandioca, sendo A) 19 indivíduos silvestres, B) 71 variedades bravas, e C) 69 variedades mansas

A) silvestres **B) bravas** **C) mansas**

Figura 22 - Gráfico de dispersão ilustrando as médias das estimativas do coeficiente de endogamia (f) para os 5.871 marcadores SNPs detectados em 159 indivíduos de mandioca, sendo A) 19 indivíduos silvestres, B) 71 variedades bravas, e C) 69 variedades mansas

Tabela 12 - Estimativas de diversidade genômica para 5.871 marcadores SNPs avaliados em 159 indivíduos de mandioca, considerando variedades de mandioca cultivada e indivíduos silvestres, variedades cultivadas ao longo dos rios, e variedades mansas e bravas ao longo dos rios. Número de indivíduos (N), número total de alelos (A), riqueza alélica média por loco (A_R), heterozigosidades observada (H_O) e esperada (H_E), e coeficiente de endogamia (f). *Significativo com intervalo de confiança de 95%

Grupos de variedades	N	A	A_R	H_O	H_E	f
Bravas	71	8.942	1,33	0,114	0,115	0,024
Mansas	69	10.274	1,33	0,120	0,109	-0,015*
Silvestres	19	8.007	1,26	0,081	0,088	0,115*
Rios						
Madeira	37	8.597	1,31	0,115	0,112	-0,002*
Negro/Branco	40	8.743	1,33	0,112	0,116	0,035*
Solimões	33	8.483	1,32	0,113	0,114	0,025
Amazonas	31	10.025	1,41	0,131	0,131	0,013
Rios (variedades)						
Madeira bravas	19	8.200	1,32	0,118	0,113	-0,018*
Madeira mansas	18	8.063	1,28	0,112	0,099	-0,076*
Negro/Branco bravas	23	8.253	1,31	0,107	0,108	0,015
Negro/Branco mansas	17	8.125	1,29	0,118	0,104	-0,070*
Solimões bravas	17	8.124	1,31	0,114	0,111	-0,011*
Solimões mansas	16	7.796	1,27	0,112	0,097	-0,095*
Amazonas bravas	19	8.154	1,33	0,124	0,120	-0,019*
Amazonas mansas	12	9.579	1,44	0,136	0,132	-0,012*

Enquanto as estimativas de divergência com base em valores de F_{ST} entre variedades de mandioca mansas e bravas havia sido de 0,005 e 0,08 para marcadores cpSSR e ncSSR, respectivamente, com base na variação genômica foi obtido o maior grau de divergência entre as variedades mansas e bravas, com $F_{ST} = 0,10$ (Tabela 13). Também foi observado um aumento no valor de divergência entre indivíduos silvestres e variedades cultivadas ($F_{ST} = 0,37$) com relação aos marcadores ncSSR ($F_{ST} = 0,28$). A diferença entre os grupos de variedades cultivadas agrupadas de acordo com os rios novamente foi baixa, mas ainda significativa ($F_{ST} = 0,013$). Ao contrário da análise com base na variação de marcadores ncSSR, não houve diferença entre grupos quando se considera a variação presente entre variedades mansas e bravas agrupadas por rio ($F_{CT} = -0,046$).

Similarmente ao observado com base na variação de marcadores ncSSR, as elevadas estimativas de divergência genômica entre variedades cultivadas e os indivíduos silvestres podem ser verificadas também através dos valores de F_{ST} par-a-par (Figura 23). Os indivíduos silvestres foram um pouco mais diferenciados das variedades mansas (0,40) do que das variedades bravas (0,38), sendo estes valores bem mais elevados do que a divergência entre as variedades cultivadas (0,10). A diferenciação entre os rios foi bem menor, variando entre 0,005 (Solimões e Negro/Branco) e 0,021 (Amazonas e Negro/Branco). A considerável divergência encontrada entre mandiocas mansas e bravas foi refletida também na comparação entre estes

grupos de variedades por rio, onde as comparações entre mandiocas mansas e bravas de diferentes rios resultaram em valores de F_{ST} mais elevados.

Tabela 13 - Análise de variância molecular (AMOVA) baseada na média de 5871 locos SNPs entre variedades mansas e bravas, entre indivíduos silvestres e variedades cultivadas de mandioca, e entre grupos de variedades de acordo com rios. Valores em negrito indicam as estatísticas correspondentes a diferenciação entre grupos (F_{ST}) em cada AMOVA. *Significativo com $p < 0,05$

Fonte de variação	Soma de quadrados	Coefficiente de variação	Porcentagem da variação (%)	Estatísticas F
Entre mansas e bravas	5.583,77	37,79	10,28	$F_{ST} = 0,103^*$
Dentro de mansas e bravas	91.078,82	329,66	89,72	
Entre silvestres e cultivadas	13.210,99	198,62	36,99	$F_{ST} = 0,369^*$
Dentro de silvestres e cultivadas	105.983,70	338,32	63,01	
Entre rios	1.988,77	4,59	1,31	$F_{ST} = 0,013^*$
Dentro de rios	94.673,82	345,17	98,69	
Entre grupos de variedades por rios	1.857,21	-15,12	-4,58	$F_{CT} = -0,046$
Entre variedades mansas e bravas dentro de rios	6.548,22	38,60	11,68	$F_{SC} = 0,112^*$
Dentro de grupos de variedades por rios	83.504,69	307,00	92,90	

$F_{ST} = 0,071^*$

Figura 23 - Estimativas de F_{ST} (diagonal inferior) e *heatmap* dos valores de F_{ST} (diagonal superior), com base em 5.871 marcadores SNPs, entre pares de: A) variedades mansas, bravas e silvestres; B) entre grupos de variedades de acordo com os rios; C) entre variedades mansas (M) e bravas (B) de acordo com rios. Valores em negrito são significativos com intervalo de confiança de 95%. As cores variam do branco (menores valores de F_{ST}) ao azul escuro (maiores valores de F_{ST})

A identificação de locos SNPs *outliers* com os aplicativos *Lositan* e *fdist2* indicou a presença de 658 locos *outliers* quando comparadas as distribuições das estimativas de F_{ST} e heterozigidades para as mandiocas mansas e bravas e indivíduos silvestres (Figura 24A). Quando considerados apenas as variedades cultivadas um total de 202 locos apresentaram comportamento *outlier* (Figura 24B).

Figura 24 - Gráficos da distribuição das estimativas de F_{ST} e heterozigosidades para detecção de locos *outliers*, com base em 5.871 marcadores SNP, obtidos com os programas *Lositan* e *fdist2*. Cada ponto preto representa um loco SNP, e as linhas vermelha e azul representam os limites superior e inferior de 99% do intervalo de confiança. Os locos que estão acima do limite superior são considerados hipoteticamente sob seleção diferencial. São considerados em A) variedades mansas, bravas e silvestres, e em B) variedades mansas e bravas

A Figura 25 ilustra a distribuição geográfica dos alelos encontrados com base nos 202 locos SNPs que apresentaram comportamento *outlier* para as variedades mansas e bravas avaliadas. As variedades de mandioca bravas apresentaram maior número total de alelos nas regiões do alto rio Negro, médio rio Amazonas e médio rio Madeira (Figura 25A), tendo a maioria das demais localidades níveis intermediários de diversidade alélica. Considerando as variedades de mandioca mansa, regiões com elevado número total de alelos estão presentes ao longo do rio Branco em Roraima, e também no baixo rio Amazonas (Figura 25B), tendo a maioria das demais localidades níveis de intermediário a baixos de diversidade alélica. A distribuição das riquezas alélica de variedades de mandioca mansas e bravas apresentaram padrões similares ao número total de alelos e não foram apresentadas. De modo geral a análise da distribuição da diversidade alélica baseada em marcadores SNPs apresentou os mesmos padrões da análise realizada com base em marcadores ncSSR (Figura 15).

Figura 25 - Mapas mostrando o número total de alelos em células de 10 minutos (aproximadamente 18 Km) aplicando o método de *circular-neighborhood* de 0,45045 graus (aproximadamente 50 Km), com base na variação genômica revelada por 202 marcadores SNPs com comportamento *outlier*. Em A) 71 variedades de mandioca brava, e em B) 69 variedades de mandioca mansa

As análises de agrupamento realizadas pelo método de *Neighbor-Joining*, bem como as de estruturação genética realizadas por meio de DAPCs, considerando todos os 5.871 locos SNPs resultaram em dendrogramas muito similares aos das análises realizadas considerando apenas os 5.213 locos SNPs neutros. Por isso apenas os resultados obtidos com os locos SNPs neutros serão apresentados. Os dendrogramas construídos para os indivíduos com base nos locos SNPs neutros apresentaram relacionamentos e topologia similares aos construídos com base nos locos SNPs *outliers* (Figura 26). Os indivíduos silvestres formam grupos consistentes na base dos dendrogramas. Em ambas as árvores duas variedades mansas proveniente do município de Oriximiná (rio Amazonas, estado do Pará) formam o grupo mais próximo dos indivíduos silvestres. Entretanto, ao contrário da topologia obtida com marcadores ncSSR (Figura 19) as mandiocas bravas formam um grupo consistente e mais próximo aos indivíduos silvestres, enquanto as mandiocas mansas formam outro grupo consistente mais interno, com a presença de algumas mandiocas bravas em ambas as árvores. Não houve formação de agrupamentos de acordo com os rios.

Figura 26 - Dendrogramas mostrando os relacionamentos entre 159 indivíduos de variedades de mandioca construído pelo método *Neighbor-Joining*, com base em distâncias de Cavalli-Sforza & Edwards (1967). São considerados em A) 5.213 marcadores SNPs com comportamento neutro, e B) 658 marcadores SNPs com comportamento *outlier*

A separação das variedades de mandiocas mansas e bravas é uma vez mais evidente quando dendrogramas são construídos com base nos grupos de variedades cultivadas ao longo dos rios (Figura 27). Novamente as árvores resultaram em topologias e relacionamentos similares considerando os locos SNPs neutros e considerando apenas os locos *outliers*. As mandiocas bravas apresentaram relacionamentos similares aos encontrados com base em marcadores ncSSR (Figura 20). Já as mandiocas mansas apresentaram relacionamentos um pouco diferentes, com as variedades dos rios Madeira e Solimões mais relacionadas entre si. Em ambos os dendrogramas os grupos de mandiocas bravas e mansas foram igualmente relacionados ao grupo de indivíduos silvestres.

A) SNPs neutros

B) SNPs outliers

Figura 27 - Dendrogramas mostrando relacionamentos entre mandiocas mansas e bravas de acordo com os rios, construído pelo método *Neighbor-Joining*, com base em distâncias de Cavalli-Sforza & Edwards (1967). São considerados em A) 5.213 marcadores SNPs com comportamento neutro, e B) 658 marcadores SNPs com comportamento *outlier*. Valores de *bootstrap* acima de 70% são apresentados

A princípio as análises de estruturação genética realizada por meio de DAPCs revelaram os mesmos padrões de relacionamento obtidos com os dendrogramas, com resultados similares para os locos SNPs neutros e para os locos SNPs *outliers* (Figuras 28 e 29). Por meio do método de *K-means* o número de agrupamentos que melhor explica os dados de variação genômica entre as variedades cultivadas e os indivíduos silvestres foi $K = 3$ para as análises realizadas com marcadores SNPs neutros e *outliers* (Figuras 28A e C). Quando os grupos de variedades são representados nas DAPCs pode-se observar uma marcante separação entre as variedades cultivadas e os indivíduos silvestres (componente principal 1, Figuras 28B e D). Em menor escala há também uma consistente separação das mandiocas mansas e bravas, com alguns indivíduos de variedades distintas mais proximamente relacionados (componente principal 2, Figuras 28B e D).

Figura 28 - Análises Discriminantes de Componentes Principais (DAPC) com marcadores SNPs para 159 indivíduos de mandioca cultivada e silvestre. Acima são apresentadas as análises com base em 5.213 marcadores SNPs com comportamento neutro, e abaixo são apresentadas as análises com base em 658 marcadores SNPs com comportamento *outlier*. Estão representados em A) e C) os agrupamentos formados pelo método de *K-means*, e em B) e D) os grupos de variedades mansas, bravas e silvestres

Quando as DAPC foram realizadas apenas com as variedades cultivadas, o método de *K-means* encontrou dois agrupamentos com base nos 5.669 marcadores SNPs neutros, e cinco agrupamentos com base nos 202 marcadores SNPs com comportamento *outlier* (Figura 29). Para os marcadores SNPs neutros a formação de grupos correspondentes às mandiocas mansas e bravas permanece, havendo a sobreposição de alguns indivíduos de ambos os grupos (Figuras 29A e B). Para os marcadores SNPs *outliers*, a separação entre grupos de variedades mansas e bravas também é evidente, porém as mandiocas bravas são subdivididas em quatro grupos (Figuras 29C e D). Quando as variedades cultivadas são representadas na DAPC baseada em SNPs *outliers* observa-se que as mandiocas bravas estão mais amplamente dispersas em relação às mandiocas mansas (Figura 29D).

Figura 29 - Análises Discriminantes de Componentes Principais (DAPC) com marcadores SNPs para 140 indivíduos de variedades de mandioca mansas e bravas. Acima são apresentadas as análises com base em 5.669 marcadores SNPs com comportamento neutro, e abaixo são apresentadas as análises com base em 202 marcadores SNPs com comportamento *outlier*. Estão representados em A) e C) os agrupamentos formados pelo método de *K-means*, e em B) e D) os grupos de variedades mansas e bravas

As diferenças entre as DAPCs realizadas com todos os marcadores SNPs neutros e apenas com os marcadores SNPs *outliers* são melhor visualizadas quando as probabilidades de atribuição dos indivíduos aos grupos são representadas em gráficos que podem ser interpretados de forma semelhante aos resultados do programa *Structure*. As análises destas probabilidades indicam a clara separação dos indivíduos silvestres e das variedades de mandioca cultivada, e uma separação não tão clara entre as variedades mansas e bravas (Figura 30A), quando considerados os marcadores SNPs neutros. A separação entre variedades cultivadas e indivíduos silvestres permanece quando considerados os 658 marcadores SNPs *outliers*, mas são formados outros dois grupos consistentes com as variedades de mandiocas mansas e bravas (Figura 30B), e que indicam a ocorrência de indivíduos com ancestralidade misturada. Quando são avaliadas apenas as variedades cultivadas com os marcadores SNPs neutros, uma vez mais ocorre mistura entre alguns indivíduos de diferentes variedades (Figura 30C). Entretanto,

quando são considerados os 202 marcadores SNPs *outliers* a separação de mandiocas mansas e bravas persiste, mas desta vez o grau de mistura é consideravelmente maior em indivíduos de variedades bravas (Figura 30D).

A) SNPs neutros

B) SNPs outliers

C) SNPs neutros

D) SNPs outliers

Figura 30 - Gráficos de barra representando as probabilidades de atribuição dos indivíduos aos agrupamentos considerados nas Análises Discriminante de Componentes Principais (DAPCs) com base em marcadores SNPs. DAPCs realizadas para variedades mansas, bravas e silvestres com base em: A) 5.213 marcadores SNPs com comportamento neutro, e B) com os 658 marcadores SNPs com comportamento *outlier*. DAPCs realizada para variedades mansas e bravas com base em: C) 5.669 marcadores SNPs com comportamento neutro, e D) com os 202 marcadores SNPs com comportamento *outlier*

Quando as DAPCs são realizadas com as 71 variedades de mandiocas bravas o método de *K-means* encontrou dois agrupamentos com base nos marcadores SNPs neutros (Figura 31A), e três agrupamentos com base nos marcadores SNPs *outliers* (Figura 31B). A interpolação espacial em mapas das probabilidades de atribuição aos agrupamentos encontrados na DAPC realizada com marcadores SNPs neutros revela que um dos agrupamentos é formado

predominantemente por variedades bravas do alto e médio rio Negro, sendo que as variedades bravas provenientes do rio Branco são um pouco menos relacionadas com este grupo. O outro agrupamento é formado por variedades bravas do alto rio Madeira e do alto rio Amazonas, sendo que as variedades das demais localidades ocorrem nos três agrupamentos. Estes resultados corroboram a interpolação realizada com as probabilidades de ancestralidade (q) de variedades bravas resultantes da análise como programa *Structure* para os marcadores ncSSR (Figura 16B). Dos três agrupamentos encontrados pela DAPC realizada com marcadores SNPs *outliers*, apenas um deles está associado às variedades bravas do médio e alto rio Amazonas, sendo as variedades das demais localidades representadas nos três agrupamentos (Figura 31B).

Figura 31 - Análises Discriminantes de Componentes principais (DAPC) e interpolação espacial das probabilidades de atribuição aos agrupamentos identificados pelas DAPCs para 71 variedades de mandiocas bravas, com base em marcadores SNPs. Mapas são mostrados para cada um dos agrupamentos. A barra de escala à direita dos mapas representa a probabilidade dos indivíduos (pontos pretos) de pertencerem ao agrupamento. Estão representadas as análises com: A) 5.669 marcadores SNPs com comportamento neutro, e B) 202 marcadores SNPs com comportamento *outlier*

Para as DAPCs realizadas com as 69 variedades de mandiocas mansas o método de *K-means* encontrou três agrupamentos com base nos marcadores SNPs neutros (Figura 32A), e quatro agrupamentos com base nos marcadores SNPs *outliers* (Figura 32B).

A) SNPs neutros

B) SNPs outliers

Figura 32 - Análises Discriminantes de Componentes principais (DAPC) e interpolação espacial das probabilidades de atribuição aos agrupamentos identificados pelas DAPCs para 69 variedades de mandiocas mansas, com base em marcadores SNPs. Mapas são mostrados para cada um dos agrupamentos. A barra de escala à direita dos mapas representa a probabilidade dos indivíduos (pontos pretos) de pertencerem ao agrupamento. Estão representadas as análises com: A) 5.669 marcadores SNPs com comportamento neutro, e B) 202 marcadores SNPs com comportamento *outlier*

Na DAPC realizada com SNPs neutros dois agrupamentos apresentam tendências de associação a regiões geográficas. O primeiro deles é formado predominantemente pelas variedades mansas do alto rio Amazonas, enquanto o segundo é formado predominantemente

por variedades mansas do médio rio Negro e do médio rio Amazonas (Figura 32A). A interpolação espacial das probabilidades de atribuição aos quatro agrupamentos encontrados na DAPC realizada com marcadores SNPs *outliers* para as variedades mansas mostra que apenas um destes agrupamentos está associado com as variedades provenientes do médio rio Madeira (Figura 32B). As demais variedades mansas estão representadas nos quatro agrupamentos.

As análises de estrutura genômica entre grupos de variedades realizadas por meio das DAPCs revelaram que, como esperado, a maior divergência genômica pode ser encontrada entre as variedades cultivadas e os indivíduos silvestres. Em seguida, as DAPCs foram capazes de recuperar uma apreciável divergência entre as variedades de mandioca mansas e bravas. Estes padrões puderam ser observados tanto nas análises com os marcadores SNPs neutros, como com marcadores SNPs *outliers*, e estão em concordância com os resultados obtidos com marcadores ncSSR. Foram identificados 97 marcadores SNPs (40 com comportamento neutro e 57 com comportamento *outlier*) que apresentaram as maiores contribuições para a variação observada nas DAPCs realizadas entre os grupos de variedades (Figuras 28 e 29). Para avaliar a base genômica destas divergências os locos do genoma de mandioca em que estes marcadores estão inseridos foram identificados, e foi verificada a associação destes locos com regiões gênicas transcritas. Dos 97 marcadores identificados, 57 estavam localizados em transcritos associados às regiões expressas do genoma parcial da mandioca *Manihot esculenta* v4.1. Destes, 22 eram SNPs neutros (Tabela 14) e 34 eram SNPs *outliers* (Tabela 15).

Entre os SNPs neutros com maior contribuição para a variação observada entre variedades cultivadas e indivíduos silvestres, os seis marcadores que estavam localizados em regiões transcritas foram responsáveis por pouco mais de 2% do total de 76,55% de variação observada no componente principal 1 da Figura 28B. Entretanto, apenas dois destes locos estavam localizados em transcritos com anotação funcional, que estão associados ao metabolismo de lipídeos e importação de proteínas para o núcleo (Tabela 14). Ainda com relação à análise da Figura 28B, dos marcadores com maior contribuição para a variação observada entre variedades de mandiocas mansas e bravas, seis locos SNPs estavam localizados em regiões transcritas, e explicaram 3,14% do total de 4,87% de variação observada no componente principal 2. Todos os transcritos nos quais estes locos SNPs estão inseridos possuem anotação funcional, e estão associados a funções moleculares e processos biológicos diversos, como ligação à moléculas orgânicas e íons inorgânicos, atividades metabólicas e transporte de moléculas. É interessante notar que o loco SNP neutro com maior contribuição para a divergência entre variedades mansas e bravas (L4098) está em uma região gênica associada à respostas de defesa (Tabela 14).

Para a DAPC realizada apenas para as variedades mansas e bravas, dez SNPs neutros estavam associados com regiões transcritas do genoma, e representaram 4,6% do total de 78,52% de variação observada no componente principal 1 da Figura 29B. Seis destes marcadores SNPs estão localizados nos mesmos transcritos ou estão entre os marcadores identificados nas análises do parágrafo anterior. Mais uma vez o loco L4098, presente em um gene associado à respostas de defesa, foi o marcador que mais contribuiu para a divergência entre variedades mansas e bravas. Destaca-se também o loco L192, localizado em uma região associada à regulação da transcrição (Tabela 14).

Com relação aos SNPs *outliers*, 19 marcadores responsáveis por 35,46% da divergência entre variedades cultivadas e indivíduos silvestres (do total de 69,2% de variação observada no componente principal 1 da Figura 28D) estavam localizados em regiões transcritas do genoma. A maioria destes SNPs está localizado em regiões gênicas associadas à processos catalíticos e transporte de moléculas orgânicas e íons. Este é o caso do loco L3263, que teve a maior contribuição para a variação observada, e está localizado em um transcrito associado ao transporte transmembrana e transporte de fosfatos (Tabela 15). Na análise da Figura 28D, também foram identificados quatro marcadores SNPs *outliers* responsáveis por 8,39% do total de 16,48% da variação observada no componente principal 2, que explica a separação entre as variedades mansas e bravas. Dois destes marcadores estão localizados em transcritos com anotação funcional, e que são associados ao metabolismo de lipídeos e atividades catalíticas (Tabela 15).

Para as análises realizadas apenas com as variedades de mandioca mansas e bravas, foram encontrados 5 marcadores SNPs *outliers* localizados em regiões transcritas. Estes foram responsáveis por 15,74% da variação entre os agrupamentos de variedades bravas (do total de 49,43% da variação observada no componente principal 1 da Figura 29C). Estes marcadores estão localizados em transcritos associados a processos diversos, dentre os quais atividades catalíticas, ligação à moléculas orgânicas e inorgânicas, e processos metabólicos. Chamam atenção os locos L1685 e L2988, associados a regiões com atividade de regulação da transcrição e respostas de defesa, respectivamente (Tabela 15). Por fim, foram encontrados 6 marcadores SNPs localizados em regiões transcritas do genoma e que explicam 23,48% da variação observada no componente principal 2 da Figura 29D, e que estavam associados à diferenciação dos grupos de variedades mansas e bravas. Entre estes, os locos L4341 e L4343 estão localizados na mesma região transcrita. De modo geral, os locos SNPs *outliers* identificados nesta última análise estão localizados em regiões gênicas associadas a ligação com moléculas orgânicas e íons inorgânicos, e processos catalíticos (Tabela 15).

Tabela 14 - Identificação, características genômicas e anotação funcional dos marcadores SNPs com comportamento neutro que apresentaram maior contribuição para a divergência de grupos de variedades de mandioca identificados pelas análises discriminantes de componentes principais (DAPCs) das Figuras 19 B e 20 B e que estavam associados à regiões transcritas do genoma parcial da mandioca (*Manihot esculenta v4.1*). São apresentados: os alelos de cada loco (SNP); a porcentagem da divergência entre grupos de variedades de mandioca explicada por cada loco nas DAPCs (%); a posição do loco SNP ao longo do genoma parcial da mandioca (*Scaffold* e *Posição*); o transcrito e a região gênica no qual o loco SNP está inserido; a anotação funcional baseada nas três classes de termos GO associada aos genes resultantes dos transcritos, depositados no banco de dados do *Joint Genome Institute* – JGI

(continua)

Locos neutros	SNP ‡	(%)	Scaffold	Posição	Transcrito	Região	Termos GO §		
							Componente celular	Função molecular	Processo biológico
Figura 28 B: silvestres x cultivadas									
L823	G / A	0,33	02815	113695	cassava4.1_019364m	éxon	SAF		
L1796	A / G	0,41	03942	217288	cassava4.1_002182m	éxon	STG		
L2161	A / G	0,41	04300	201701	cassava4.1_017747m	íntron	STG		
L3813	T / C	0,32	07520	966964	cassava4.1_006366m	éxon	SAF		
L4576	T / A	0,38	09120	498364	cassava4.1_010094m	éxon		GO:0016788 atividade de hidrolase em ligações éster	GO:0006629 metabolismo de lipídeos
L4833	A / G	0,16	09842	1612	cassava4.1_024283m	íntron	GO:0005634 núcleo GO:0005737 citoplasma	GO:0016787 atividade de hidrolase GO:0005515 ligação à proteínas GO:0008565 atividade de transporte de proteínas	GO:0006606 importação de proteínas para o núcleo
Figura 28 B: mansas x bravas									
L344	A / C	0,52	01481	166140	cassava4.1_023181m	éxon		GO:0004553 atividade de hidrolase, em compostos O-glicosil	GO:0005975 metabolismo de carboidratos
L392	A / T, C	0,27	01551	966171	cassava4.1_008988m	5' UTR	GO:0016020 membrana	GO:0030246 ligação à carboidratos GO:0046873 atividade de transporte transmembrana de íons metálicos	GO:0030001 transporte de íons metálicos GO:0055085 transporte transmembrana
L994	C / T	0,61	02960	640829	cassava4.1_010565m	éxon		GO:0005515 ligação à proteínas GO:0008270 ligação a íons Zn	GO:0006952 resposta de defesa
L4098	G / A	0,66	08020	5166	cassava4.1_001198m	éxon		GO:0043531 ligação a ADP	
L4404	G / T	0,55	08542	551340	cassava4.1_001856m	éxon	GO:0016020 membrana	GO:0000166 ligação a nucleotídeos GO:0016887 atividade ATPase GO:0005524 ligação a ATP GO:0017111 atividade nucleosideo-trifosfatase	
L5691	T / G	0,53	12317	72049	cassava4.1_011415m	éxon		GO:0003676 ligação à ácidos nucléicos GO:0004518 atividade de nuclease	

Tabela 14 - Identificação, características genômicas e anotação funcional dos marcadores SNPs com comportamento neutro que apresentaram maior contribuição para a divergência de grupos de variedades de mandioca identificados pelas análises discriminantes de componentes principais (DAPCs) das Figuras 19 B e 20 B e que estavam associados à regiões transcritas do genoma parcial da mandioca (*Manihot esculenta v4.1*). São apresentados: os alelos de cada loco (SNP); a porcentagem da divergência entre grupos de variedades de mandioca explicada por cada loco nas DAPCs (%); a posição do loco SNP ao longo do genoma parcial da mandioca (*Scaffold* e *Posição*); o transcrito e a região gênica no qual o loco SNP está inserido; a anotação funcional baseada nas três classes de termos GO associada aos genes resultantes dos transcritos, depositados no banco de dados do *Joint Genome Institute* – JGI

							(conclusão)		
							Termos GO §		
Locos neutros	SNP ‡	(%)	Scaffold	Posição	Transcrito	Região	Componente celular	Função molecular	Processo biológico
Figura 29 B: mansas x bravas									
L192	C / A	0,41	00823	384368	cassava4.1_025203m	íntron		GO:0003700 atividade de fator de transcrição com ligação à sequências específicas de DNA GO:0043565 ligação à sequências específicas de DNA	GO:0006355 regulação de transcrição, DNA-dependente
L344	A / C	0,4	01481	166140	cassava4.1_023181m	éxon		GO:0004553 atividade de hidrolase, em compostos O-glicosil GO:0030246 ligação à carboidratos	GO:0005975 metabolismo de carboidratos
L390	T / G	0,47	01551	966146	cassava4.1_008988m	5' UTR	GO:0016020 membrana	GO:0046873 atividade de transporte transmembrana de íons metálicos	GO:0030001 transporte de íons metálicos GO:0055085 transporte transmembrana
L392	A / T, C	0,28	01551	966171	cassava4.1_008988m	5' UTR	GO:0016020 membrana	GO:0046873 atividade de transporte transmembrana de íons metálicos	GO:0030001 transporte de íons metálicos GO:0055085 transporte transmembrana
L994	C / T	0,42	02960	640829	cassava4.1_010565m	éxon		GO:0005515 ligação à proteínas GO:0008270 ligação a íons Zn GO:0005525 ligação a GTP	
L2255	G / C	0,42	04457	1230178	cassava4.1_010161m	íntron	SAF		
L2541	T / C	0,46	05214	1274892	cassava4.1_029560m	éxon			
L4098	G / A	0,76	08020	5166	cassava4.1_001198m	éxon		GO:0043531 ligação a ADP	GO:0006952 resposta de defesa
L4404	G / T	0,53	08542	551340	cassava4.1_001856m	éxon	GO:0016020 membrana	GO:0000166 ligação a nucleotídeos GO:0016887 atividade ATPase GO:0005524 ligação a ATP GO:0017111 atividade nucleosídeo-trifosfatase	
L5450	G / A	0,45	11581	798966	cassava4.1_033639m	éxon	SAF		

‡ O alelo de referência (presente no genoma parcial da mandioca *Manihot esculenta v4.1*) é indicado antes da barra, e as variantes detectadas após a barra.

§ SAF = Sem Anotação Funcional, STG = Sem Termos GO

Tabela 15 - Identificação, características genômicas e anotação funcional dos marcadores SNPs com comportamento *outlier* que apresentaram maior contribuição para a divergência de grupos de variedades de mandioca identificados pelas análises discriminantes de componentes principais (DAPCs) das Figuras 19 D e 20 D e que estavam associados à regiões transcritas do genoma parcial da mandioca (*Manihot esculenta v4.1*). São apresentados: os alelos de cada loco (SNP); a porcentagem da divergência entre grupos de variedades de mandioca explicada por cada loco nas DAPCs (%); a posição do loco SNP ao longo do genoma parcial da mandioca (*Scaffold* e *Posição*); o transcrito e a região gênica no qual o loco SNP está inserido; a anotação funcional baseada nas três classes de termos GO associada aos genes resultantes dos transcritos, depositados no banco de dados do *Joint Genome Institute* – JGI

(continua)									
Locos <i>outliers</i>	SNP ‡	(%)	Scaffold	Posição	Transcrito	Região	Termos GO §		
							Componente celular	Função molecular	Processo biológico
Figura 28 D: silvestres x cultivadas									
L674	A / G	1,75	02431	503319	cassava4.1_005057m	3' UTR		GO:0005524 ligação à ATP GO:0051082 ligação à proteínas não dobradas	GO:0044267 processo metabólico de proteínas celulares GO:0006457 dobramento de proteínas
L1406	T / C	1,93	03581	1492238	cassava4.1_003646m	5' UTR	STG		
L1723	T / A	2,07	03802	304604	cassava4.1_011094m	íntron		GO:0016787 atividade de hidrolase	
L1794	G / A	1,83	03942	217260	cassava4.1_002182m	éxon	STG		
L2090	G / A	1,7	04233	201420	cassava4.1_013257m	íntron	STG		
L2194	G / A	1,89	04374	108122	cassava4.1_011709m	éxon	STG		
L2210	G / C	1,76	04457	116997	cassava4.1_025451m	íntron	GO:0005634 núcleo GO:0005737 citoplasma	GO:0005515 ligação à proteínas GO:0008565 atividade de transporte de proteínas	GO:0006606 importação de proteínas para o núcleo
L2440	C / T	1,97	04957	24399	cassava4.1_007951m	íntron	GO:0016021 integral à membrana	GO:0019829 atividade de transporte de íons de ATPase	GO:0006812 transporte de cátions
L2895	C / G	1,68	06327	484023	cassava4.1_001909m	íntron		GO:0005515 ligação à proteínas GO:0008270 ligação a íons Zn	
L3262	G / T	2,08	06715	60660	cassava4.1_005227m	éxon	GO:0016021 integral à membrana	GO:0005315 atividade de transporte transmembrana de fosfatos inorgânicos GO:0022857 atividade de transporte transmembrana	GO:0006817 transporte de íons fosfato GO:0055085 transporte transmembrana
L3815	G / T	1,82	07520	967073	cassava4.1_006366m	íntron	SAF		
L3816	T / A	1,82	07520	967115	cassava4.1_006366m	íntron	SAF		
L3895	A / G	1,91	07528	273224	cassava4.1_005422m	éxon		GO:0016705 atividade de oxireductase, agindo em doadores pareados, com incorporação ou redução de oxigênio molecular GO:0020037 ligação a grupo heme GO:0005506 ligação a íons Fe	GO:0055114 processos de oxi-redução
L4077	G / A	1,78	07991	653689	cassava4.1_001103m	íntron			GO:0006397 processamento de mRNA

Tabela 15 - Identificação, características genômicas e anotação funcional dos marcadores SNPs com comportamento *outlier* que apresentaram maior contribuição para a divergência de grupos de variedades de mandioca identificados pelas análises discriminantes de componentes principais (DAPCs) das Figuras 19 D e 20 D e que estavam associados à regiões transcritas do genoma parcial da mandioca (*Manihot esculenta v4.1*). São apresentados: os alelos de cada loco (SNP); a porcentagem da divergência entre grupos de variedades de mandioca explicada por cada loco nas DAPCs (%); a posição do loco SNP ao longo do genoma parcial da mandioca (*Scaffold* e *Posição*); o transcrito e a região gênica no qual o loco SNP está inserido; a anotação funcional baseada nas três classes de termos GO associada aos genes resultantes dos transcritos, depositados no banco de dados do *Joint Genome Institute* – JGI

(continuação)									
Locos <i>outliers</i>	SNP ‡	(%)	Scaffold	Posição	Transcrito	Região	Termos GO §		
							Componente celular	Função molecular	Processo biológico
Figura 28 D: silvestres x cultivadas									
L4350	G / A	1,86	08494	100007	cassava4.1_019368m	íntron	GO:0005840 ribossomo GO:0005622 intracelular	GO:0003735 constituinte estrutural de ribossomos	GO:0006412 tradução
L4729	T / C	2,04	09501	661186	cassava4.1_010397m	3' UTR		GO:0046983 atividade de dimerização de proteínas	
L5282	A / T	1,96	11179	131861	cassava4.1_007673m	íntron	GO:0000015 complexo fosfopiruvato hidratase	GO:0004634 atividade de fosfopiruvato hidratase	GO:0006096 glicólise
L5621	T / A	1,73	12004	557337	cassava4.1_031073m	éxon		GO:0000287 ligação a íons Mg GO:0004185 atividade de serino-carboxipeptidase	GO:0006508 proteólise
L5874	T / C	1,88	12865	955441	cassava4.1_021670m	éxon	STG		
Figura 28 D: mansas x bravas									
L1315	T / C	2,25	03410	78677	cassava4.1_022811m	éxon	SAF		
L1316	G / A	2,25	03410	78697	cassava4.1_022811m	éxon	SAF		
L1399	C / T	2,17	03581	1007433	cassava4.1_007605m	éxon		GO:0020037 ligação ao grupo heme	GO:0006629 processos metabólicos de lipídeos
L2992	T / G	1,72	06512	1763159	cassava4.1_002690m	éxon	GO:0016020 membrana	GO:0000166 ligação a nucleotídeos GO:0016887 atividade de ATPase GO:0005524 ligação à ATP GO:0017111 atividade de nucleosídeo-trifosfatase	
Figura 29 D: grupos de bravas									
L212	G / A	2,92	00847	999155	cassava4.1_005553m	éxon		GO:0016491 atividade de oxiredutase	GO:0008152 processos metabólicos
L1661	G / T	2,79	03750	462305	cassava4.1_025627m	éxon		GO:0004674 atividade de serina/treonina proteíno-quinase GO:0004713 atividade de tirosina proteino-quinase GO:0005515 ligação à proteínas GO:0005524 ligação à ATP GO:0004672 atividade de proteíno-quinase	GO:0006468 fosforilação de proteínas

Tabela 15 - Identificação, características genômicas e anotação funcional dos marcadores SNPs com comportamento *outlier* que apresentaram maior contribuição para a divergência de grupos de variedades de mandioca identificados pelas análises discriminantes de componentes principais (DAPCs) das Figuras 19 D e 20 D e que estavam associados à regiões transcritas do genoma parcial da mandioca (*Manihot esculenta* v4.1). São apresentados: os alelos de cada loco (SNP); a porcentagem da divergência entre grupos de variedades de mandioca explicada por cada loco nas DAPCs (%); a posição do loco SNP ao longo do genoma parcial da mandioca (*Scaffold* e *Posição*); o transcrito e a região gênica no qual o loco SNP está inserido; a anotação funcional baseada nas três classes de termos GO associada aos genes resultantes dos transcritos, depositados no banco de dados do *Joint Genome Institute* – JGI

										(conclusão)
Locos outliers	SNP ‡	(%)	Scaffold	Posição	Transcrito	Região	Termos GO §			
							Componente celular	Função molecular	Processo biológico	
Figura 29 D: grupos de bravas										
L1685	T / C	3,16	03750	868433	cassava4.1_029170m	éxon		GO:0003677 ligação ao DNA	GO:0006355 regulação da transcrição, DNA-dependente GO:0006952 resposta de defesa	
L2988	G / C	2,83	06512	1609996	cassava4.1_031835m	éxon	GO:0016021 integral à membrana			
L4587	C / A	4,04	09151	316844	cassava4.1_000437m	íntron		GO:0005515 ligação à proteínas		
Figura 29 D: mansas x bravas										
L4341	A / G	3,19	08485	192026	cassava4.1_000923m	íntron		GO:0005515 ligação à proteínas	GO:0006468 fosforilação de proteínas	
L4343	A / G	3,19	08485	192040	cassava4.1_000923m	íntron		GO:0005515 ligação à proteínas		
L4405	C / T	3,76	08542	689599	cassava4.1_029823m	éxon		GO:0004674 atividade de serina/treonina proteíno-quinase		
								GO:0004713 atividade de proteíno-quinase		
								GO:0005524 ligação à ATP		
								GO:0005509 ligação a íons Ca		
								GO:0004672 atividade de proteíno-quinase		
L5083	T / C	3,25	10563	110941	cassava4.1_032052m	éxon	STG			
L5418	A / C	3,58	11495	431949	cassava4.1_018317m	íntron		GO:0016881 atividade de aminoácido-ligase		
L5419	A / G	6,51	11495	444701	cassava4.1_026455m	éxon		GO:0008270 ligação a íons Zn		
								GO:0003676 ligação à ácidos nucleicos		

‡ O alelo de referência (presente no genoma parcial da mandioca *Manihot esculenta* v4.1) é indicado antes da barra, e as variantes detectadas após a barra.

§ SAF = Sem Anotação Funcional, STG = Sem Termos GO.

5 DISCUSSÃO

5.1 Diversidades haplotípica, genética e genômica de variedades de mandioca

Os quatro locos cpSSR polimórficos utilizados neste estudo apresentaram um total de 11 alelos, que quando avaliados em conjunto resultaram em 11 haplótipos cloroplastidiais. O baixo número de marcadores polimórficos encontrados na avaliação de variedades cultivadas e silvestres de mandioca neste estudo segue a tendência da maioria dos trabalhos realizados com marcadores cpSSR. Isto pode ser justificado pela utilização de iniciadores “universais”, tais como os descritos por Weising e Gardner (1999) e por Chung e Staub (2003), que foram utilizados neste estudo. Apesar destes locos “universais” poderem ser amplificados em uma grande variedade de espécies de plantas, não necessariamente revelarão elevados níveis de polimorfismo intraespecífico (PROVAN; POWELL; HOLLINGSWORTH, 2001). Uma grande variação nos níveis de polimorfismo em estudos envolvendo marcadores cloroplastidiais “universais” em espécies cultivadas pode ser observada na literatura. Em alguns casos poucos locos cpSSR apresentam polimorfismo, e revelam um baixo número de haplótipos. Boccacci e Botta (2009) avaliando germoplasma de aveleira (*Corylus avellana*), encontraram polimorfismo em quatro locos cpSSR que resultaram em apenas quatro haplótipos cloroplastidiais. Castro e colaboradores (2013) utilizaram o polimorfismo de três locos cpSSR, que permitiram a identificação de quatro haplótipos em cultivares portuguesas de videira (*Vitis vinifera*). Por outro lado, uma grande quantidade de variação também pode ser encontrada. Angioi e colaboradores (2010) encontraram 101 haplótipos a partir da avaliação de seis locos cpSSR em variedades europeias de feijão comum (*Phaseolus vulgaris*). Com relação aos índices de diversidade haplotípica, pode-se afirmar que os resultados obtidos neste trabalho revelam um nível intermediário de diversidade haplotípica, quando comparado com estudos realizados em outras espécies cultivadas, como, por exemplo, os trabalhos supracitados.

Com relação aos marcadores ncSSR, o elevado grau de polimorfismo encontrado, com média de 9 alelos por loco ($\bar{A}$) e heterozigosidade observada (H_O) de 0,616, corrobora os níveis de diversidade genética encontrados para mandioca em trabalhos anteriores. Dentre os estudos mais recentes, Alves-Pereira e colaboradores (2011) observaram médias de $\bar{A} = 5,6$ e $H_O = 0.576$ em variedades de mandioca do médio rio Madeira. Mühlen e colaboradores (2013) encontraram médias de $\bar{A} = 9,9$ e $H_O = 0,595$, avaliando variedades de todo o Brasil, a maioria provenientes da Amazônia. Estes elevados níveis de diversidade genética são esperados para marcadores ncSSR, sendo de fato uma das vantagens da utilização deste tipo de marcador (POWELL;

MACHRAY; PROVAN, 1996).

Os marcadores SNPs avaliados em 159 variedades de mandioca apresentaram níveis de heterozigosidade um pouco acima de 11%. Os valores de diversidade genômica observados neste estudo são comparáveis a outros trabalhos realizados com marcadores SNPs. Pegadaraju e colaboradores (2013) encontraram valores médios de heterozigosidade similares (10,8%) em girassol (*Helianthus annuus* L.). De outra forma, valores de heterozigosidade mais elevados foram observados, por exemplo, para coníferas (~36% para *Pinus contorta* e *Picea glauca*) por Chen et al. (2013), e videiras (~31% para *Vitis vinifera* ssp. *vinifera* e ssp. *sylvestris*) por Emanuelli et al. (2013). Menor variação foi encontrada neste trabalho em comparação com estudos anteriores que avaliaram a diversidade genômica de variedades de mandioca com marcadores SNPs. Peña-Venegas et al. (2014) encontraram heterozigosidade média de 39% avaliando 173 variedades colombianas de mandioca com base em 93 marcadores SNPs. Oliveira et al. (2014) observaram heterozigosidade média de 32% em 1.280 acessos do banco de germoplasma brasileiro de mandioca, utilizando 402 marcadores SNPs. Esta diferença provavelmente se deve ao fato destes autores terem selecionado locos com maior variação a partir do estudo que identificou e validou os marcadores SNPs em questão (FERGUSON et al., 2012). Apesar dos objetivos distintos, a cobertura média de 535X dos 5.871 marcadores SNPs encontrados neste estudo é muito maior do que, por exemplo, a encontrada por Pegadaraju et al. (2013), que consideram como confiáveis os marcadores SNPs com cobertura $\geq 4X$. De acordo com Nielsen et al. (2005), os genótipos obtidos a partir de métodos que utilizam redução da complexidade do genoma, tais como RAD-seq, podem ser aplicados na estimativa de parâmetros de genética de populações sem preocupações com vieses. A elevada cobertura média dos marcadores SNPs avaliados neste estudo certamente contribuem para o aumento da acurácia e precisão das estimativas dos parâmetros genéticos, como sugerido por Luikart et al. (2003).

Pôde-se observar com os três tipos de marcadores utilizados neste estudo que as variedades cultivadas, quando comparados com os indivíduos silvestres, apresentam níveis mais elevados de diversidade haplotípica ($N_H = 11$ versus 2; $H_{Ehap} = 0,383$ versus 0,071), genética ($A_R = 8,28$ versus 4,28; $H_O = 0,626$ versus 0,466) e genômica ($H_O = 0,117$ versus 0,081). Isto está provavelmente associado à amostragem em apenas uma pequena fração da área de ocorrência natural da espécie silvestre (LEOTÁRD et al., 2009). Enquanto os indivíduos silvestres foram amostrados em três municípios, as variedades cultivadas foram amostradas em 43 municípios. Dessa forma, não foi possível encontrar o padrão esperado de que a variação observada em variedades de mandioca cultivada seja uma subamostra da variação genética

observada no parente silvestre (LEOTÁRD et al., 2009; OLSEN; SCHAAL, 2001). Além disso, a seleção para diferentes características de interesse de acordo com as preferências de agricultores de diferentes regiões pode contribuir para o aumento a diversidade genética das populações cultivadas em relação ao parente silvestre (MILLER; SCHAAL, 2005). Os indivíduos silvestres apresentaram também coeficientes de endogamia maiores do que as variedades de mandiocas mansas e bravas para ncSSR ($f = 0,122$) e SNPs ($f = 0,115$). Estes resultados corroboram o estudo de Olsen e Schaal (2001) que encontrou déficit de heterozigotos nas 27 populações (f médio = 0,13) de *Manihot esculenta* ssp. *flabellifolia* avaliadas com cinco locos ncSSR. Como discutido por estes autores, estes resultados são esperados ao se considerar a história de vida da espécie. Apesar das inflorescências apresentarem metandria (a antese das flores masculinas precede a antese das flores femininas), não existem sistemas de auto-incompatibilidade conhecidos em *Manihot* (OLSEN; SCHAAL, 2001). Além disso, o mecanismo de dispersão de sementes por meio da explosão de frutos deiscentes não garante o fluxo gênico entre longas distâncias. Estas características podem aumentar a frequência de fecundação entre indivíduos aparentados, contribuindo para a endogamia observada em populações de *M. esculenta* ssp. *flabellifolia*.

Os elevados níveis de diversidade genética encontrados nas variedades de mandioca cultivadas podem estar associados à diferentes fatores, tais como a produção em diferentes ambientes com adoção de diferentes estratégias de produção. Um exemplo comum é o manejo de diferentes conjuntos de variedades de mandioca em diferentes tipos de solo da terra firme e ao longo das várzeas dos rios amazônicos de água barrenta (FRASER, 2010; FRASER et al., 2012). Dentre outros aspectos, diferentes conjuntos de variedades podem ser cultivados em solos distintos dependendo da percepção dos agricultores sobre a fertilidade do solo. Em solos mais férteis as variedades de mandioca podem ser cultivadas mais intensivamente, com maior número de ciclos de plantio em uma mesma roça e menor período de pousio. Por outro lado, em solos de menor fertilidade as variedades de mandioca são comumente cultivadas mais extensivamente, com menor número de ciclos de plantio e maior período de pousio (FRASER; CLEMENT, 2008). Diferentes zonas dentro das várzeas, sujeitas a regimes variáveis de inundação, também abrem a oportunidade para os agricultores manejarem diferentes conjuntos de variedades e selecionarem diferentes características de interesse (FRASER et al., 2012). Aliado a isto, os múltiplos usos da mandioca (ELIAS; RIVAL; McKEY, 2000) e objetivos dos agricultores em cada região também podem colaborar para a seleção de características variáveis, resultando desta forma na manutenção da diversidade genética.

Outros aspectos importantes do manejo das variedades por parte de agricultores tradicionais, que possuem consequências genéticas interessantes, também podem colaborar para a manutenção da variação genética e genômica observada em um cultivo que é propagado vegetativamente. Entre estes aspectos está o fluxo gênico resultante da avaliação e incorporação de plantas oriundas de reprodução sexual nos conjuntos de variedades clonais manejados (DUPUTIE et al., 2009; MARTINS, 2001). De forma consciente ou inconsciente as plântulas oriundas de sementes não são retiradas das roças, de modo que elas podem ser avaliadas pelos agricultores na época da colheita (McKEY et al., 2010a, 2012; PUJOL et al., 2005). Caso estas plantas apresentem características de interesse elas podem ser incorporadas às variedades existentes, ou mesmo dar origem à novas variedades (McKEY et al., 2010a; PERONI, 1998). Se as sementes que deram origem às plantas incorporadas forem originárias do cruzamento entre plantas com genótipos diferentes, este processo irá colaborar para a manutenção da diversidade intra-varietal e intra-específica do cultivo.

Outro fator que pode estar associado com a variação observada em mandiocas mansas e bravas é o fluxo genotípico resultante das redes de trocas e intercâmbio de variedades (OLIVEIRA, 2008; SALICK; CELLINESE; KNAPP, 1997). Redes de intercâmbio de variedades são bem documentadas também em outros cultivos associados ao manejo tradicional, como por exemplo, entre agricultores de milho (*Zea mays* ssp. *mays*) no México (LOUETTE; CHARRIER; BERTHAUD, 1997), e de batatas (*Solanum tuberosum*) na região andina (BRUSH; CARNEY, HUAMÁN, 1981). Estas redes são características do cultivo tradicional de mandioca realizado na região amazônica, e envolvem agricultores de uma mesma ou diferentes comunidades, podendo se estender até mesmo entre grandes distâncias (BOSTER, 1986; COOMES, 2010; DYER; GONZÁLEZ; LOPERA, 2011; EMPERAIRE, 2002). Elias, Rival e McKey (2000) consideram o intercâmbio de variedades de mandioca entre agricultores, por meio de estacas a serem propagadas, uma pressão seletiva importante e responsável por parte da diversidade observada nas variedades cultivadas. As redes de intercâmbio possibilitam aos agricultores a obtenção de novas variedades, recuperação de variedades perdidas, e recomposição de estoques clonais (DELÊTRE; McKEY; HODKINSON, 2011). Desta forma, estas redes dinâmicas possuem papel importante na subsistência dos agricultores tradicionais (COOMES, 2010) e na evolução do cultivo (DYER; GONZÁLEZ; LOPERA, 2011). O fluxo genotípico promovido pelas redes de intercâmbio pode resultar também em fluxo gênico entre variedades selecionadas em diferentes ambientes e com diferentes características. Isto pode acontecer caso sejam produzidas sementes resultantes do cruzamento entre variedades

provenientes de diferentes regiões que podem originar plantas que sejam avaliadas para incorporação pelos agricultores.

A grande proporção de heterozigotos observada nas variedades mansas e bravas corrobora Pujol e McKey (2006), que encontraram uma correlação entre tamanho da planta, heterozigosidade e sobrevivência em roças de mandioca. Plantas oriundas de reprodução sexual que apresentam tamanhos maiores durante sua fase de crescimento possuem maior chance de serem distinguidas pelos agricultores quando do processo de limpeza e retirada de plantas daninhas das roças. Foi verificado que os indivíduos de maior tamanho apresentam maior heterozigosidade e que possuem maior chance de se desenvolverem até a época de colheita. Assim, mais um aspecto do manejo das variedades praticado por agricultores tradicionais pode estar colaborando para a manutenção da diversidade genética e genômica observada em variedades de mandioca de diferentes localidades da Amazônia.

A maioria dos aspectos do manejo tradicional de variedades de mandioca apresentados nos parágrafos anteriores colaboram para a manutenção da diversidade genética intra-varietal do cultivo. Além destes fatores, processos genéticos podem estar diretamente relacionados com o aumento da diversidade genética observada em variedades de um cultivo propagado vegetativamente. Eventos de hibridização podem ter papel importante no aumento da diversidade genética e fenotípica em variedades cultivadas de mandioca mansas e bravas (ALLEM, 2002). No caso da hibridização intraespecífica, novos alelos podem ser incluídos dentro de variedades policlonais, mas não ocorre aumento da diversidade dentro do cultivo. De outra forma, novos alelos podem ser incorporados por meio de hibridização entre variedades cultivadas de mandioca e espécies silvestres (LADIZINSKY, 1998). Além disso, existe a possibilidade da introdução de novos alelos nas variedades de mandioca por mutação somática que eventualmente ocorrem em partes da planta que são utilizadas como propágulos (McKEY et al., 2010a). Mesmo que as taxas de mutação dos marcadores utilizados sejam pequenas, da ordem de 10^{-2} a 10^{-6} para ncSSR, e de 10^{-8} a 10^{-9} para SNPs (FREELAND, 2005), parte da atual diversidade genética apresentada por variedades de mandioca pode ser atribuída a mutações, dada a antiguidade da domesticação (cerca de 10 mil anos antes do presente, OLSEN; SCHAAL, 1999) e dispersão do cultivo.

Com base na variação observada em ncSSRs existe a tendência de que as variedades mansas apresentem maior heterozigosidade ($H_o = 0,691$ versus $0,573$) e estimativas de endogamia ($f = -0,157$ versus $0,065$) mais baixas do que as variedades bravas, apesar de riquezas alélicas similares para estes grupos. Isto foi menos evidente para os marcadores SNPs, porém

o número de alelos detectados com este marcador foi bem maior para as variedades mansas (10.274) do que para as variedades bravas (8.942). Apesar de recentemente Bradburry e colaboradores (2013) não terem detectado este padrão analisando variedades mansas e bravas de diferentes países, trabalhos realizados com variedades brasileiras também observaram maiores heterozigosidades em mandiocas mansas (ELIAS et al., 2004; MÜHLEN; MARTINS; ANDO, 2000; PERONI; KAGEYAMA; BEGOSSI, 2007). Entretanto, estes autores não fazem comentários sobre o assunto. Uma possível explicação pode estar relacionada à hipótese de Arroyo-Kalin (2010). O autor sugere que a partir do evento inicial de domesticação da mandioca, o desenvolvimento de variedades mansas ocorreu anteriormente ao desenvolvimento de variedades bravas. Este cenário poderia implicar em uma diferença temporal entre as dispersões dos grupos de variedades, sendo que a maior diversidade alélica observada nas mandiocas mansas poderia ser uma assinatura de um maior período de diversificação a partir de seu surgimento.

As análises da distribuição geográfica da diversidade alélica sugerem que as variedades mansas e bravas possuem diferentes áreas que concentram maior diversidade genética (Figura 15). Este padrão explica a relativa uniformidade da distribuição dos alelos observada para o conjunto inteiro de 596 indivíduos avaliados com ncSSR. Enquanto as mandiocas bravas apresentam o maior nível de variação no médio e baixo rio Madeira e no médio e alto rio Negro, as mandiocas mansas apresentam maior variação ao redor do rio Branco. Apesar do esforço para se implementar uma amostragem uniforme, as áreas com maior número de alelos podem simplesmente refletir o maior número de variedades amostradas (Tabela 1). A aplicação da rarefação permite que pelo menos parte deste viés de amostragem seja corrigido (VanZONNEVELD et al., 2012). Mesmo após a restrição do número de células nos mapas após a rarefação, as mandiocas mansas e bravas continuam apresentando regiões de maior diversidade alélica diferentes. Para ambos os grupos de variedades, existem áreas com grande diversidade alélica em regiões próximas ao local de origem do cultivo (médio e baixo rio Madeira para as mandiocas bravas, e região do rio Guaporé para as mandiocas mansas), no sudoeste da bacia amazônica (OLSEN; SCHAAL, 1999). As análises ainda indicam uma maior diversidade alélica de variedades mansas no médio e alto rio Negro, em oposição ao mapa ilustrado na Figura 4 (ARROYO-KALIN, 2010; EMPERAIRE, 2002), que sugere que as mandiocas bravas são mais diversificadas nesta região. Estes resultados corroboram Mühlen et al. (2013), que ao avaliarem variedades de mandioca provenientes de diferentes regiões brasileiras encontraram as maiores médias do número de alelos na região do alto e médio rio

Negro. Apesar disso, cabe ressaltar que o mapa de ocorrência e abundância de variedades de mandioca representado na Figura 4 apresenta muitas lacunas. Uma delas está na escassez de informações sobre variedades ao longo do rio Madeira e seus tributários. Esta foi a região que apresentou maior diversidade alélica para as mandiocas bravas, e também inclui a região do rio Guaporé, que apresentou elevada diversidade alélica para as mandiocas mansas.

5.2 Padrões de estruturação haplotípica, genética e genômica, e agrupamentos de variedades de mandioca

Os resultados de estruturação haplotípica, genética e genômica indicam uma clara divergência genética entre os indivíduos silvestres de *M. esculenta* ssp. *flabellifolia* e as variedades cultivadas de mandioca. Esta divergência pode ser visualizada por meio das estimativas de F_{ST} par-a-par, AMOVAs ($F_{ST}(\text{cpSSR}) = 0,78$; $F_{ST}(\text{SNPs}) = 0,37$; $F_{ST}(\text{ncSSR}) = 0,28$), nas análises de estruturação genética Bayesiana e multivariada, e também nas análises dos relacionamentos entre os grupos de variedades realizadas por meio da construção de dendrogramas. Apesar de marcantes estes resultados são esperados, visto que sob cultivo as plantas domesticadas possuem evolução contínua em resposta à seleção humana (e também outros fatores evolutivos), e com isso há a tendência de crescente diferenciação genética entre as populações domesticadas e o(s) ancestral(is) silvestre(s) (LADIZINSKY, 1998; MILLER; SCHAAL, 2005; PICKERSGILL, 2007). Estes elevados valores de divergência entre variedades cultivadas e indivíduos silvestres também podem estar relacionados à antiguidade da domesticação da mandioca.

Além da grande divergência encontrada entre os indivíduos silvestres e as variedades cultivadas, as análises de estruturação e agrupamento sugerem ainda que variedades mansas e bravas são igualmente relacionadas aos indivíduos silvestres. O haplótipo mais comum entre indivíduos silvestres (B, Figuras 11 e 12) foi compartilhado por variedades mansas e bravas. Além disso, os haplótipos que divergem do haplótipo B com menor número de mutações também ocorrem em variedades mansas ou bravas. De acordo com Miller e Schaal (2005), o compartilhamento de haplótipos cloroplastidiais pode ser atribuído ao compartilhamento de linhagens ancestrais (*incomplete lineage sorting*), à hibridização, ou ambos. Os resultados observados podem ser melhor explicados pela ancestralidade comum entre variedades mansas e bravas e os indivíduos de *M. esculenta* ssp. *flabellifolia* (OLSEN; SCHAAL, 1999). A baixíssima divergência encontrada entre as linhagens matrilineares de variedades mansas e

bravas ($F_{ST} = 0,005$), a sobreposição destes grupos de variedades na DAPC, e o compartilhamento dos haplótipos mais frequentes e de ambos haplótipos silvestres também podem ser explicados pela ancestralidade comum. Apesar de ser difícil distinguir ancestralidade comum de homoplasia em marcadores cpSSR (PROVAN; POWELL; HOLLINGSWORTH, 2001), as análises de relacionamentos entre linhagens matrilineares foram corroboradas pelas análises realizadas com os outros marcadores. Houve uma boa concordância entre os agrupamentos encontrados na análise Bayesiana de estruturação genética realizada com ncSSR e as variedades mansas e bravas (Figura 16A). Nesta análise os indivíduos silvestres apresentaram probabilidades semelhantes de pertencerem aos dois agrupamentos (Figura 16A). Nos dendrogramas construídos com base em ncSSR e SNPs as variedades mansas e bravas de diferentes rios formam dois grupos distintos e consistentes, e estão igualmente relacionados com *M. esculenta* ssp. *flabellifolia* (Figuras 20 e 27). Por outro lado, a hibridização interespecífica pode ter sido a responsável pela presença do haplótipo B (mais comum entre indivíduos silvestres) em variedades de mandioca cultivada em Roraima (Figura 12), apesar de não terem sido amostrados indivíduos silvestres nesta região.

Apesar da evidência genética apontar a espécie *M. esculenta* ssp. *flabellifolia* como parente silvestre mais próximo da mandioca cultivada (OLSEN; SCHAAL, 1999), ainda há discussão acerca da ordem dos eventos de domesticação de variedades mansas e bravas. McKey e Beckerman (1993) resumem as quatro hipóteses principais: (i) mandiocas silvestres mansas deram origem a variedades mansas e a partir destas foram criadas variedades bravas; (ii) mandiocas silvestres bravas deram origem a variedades bravas e a partir destas foram criadas variedades mansas; (iii) mandiocas silvestres mansas deram origem a variedades mansas independentemente de mandiocas silvestres bravas que originaram variedades bravas; (iv) mandioca silvestre possivelmente com toxidez intermediária que originaram variedades mansas e bravas simultaneamente.

Ainda que permaneça uma questão muito especulativa, Arroyo-Kalin (2010) propôs que as variedades mansas foram domesticadas primeiro e posteriormente - em decorrência da intensificação da agricultura dos povos amazônicos - as variedades bravas surgiram. Desta forma, uma quinta hipótese pode ser adicionada: (v) mandiocas silvestres possivelmente com toxidez intermediária que deram origem às variedades mansas, das quais as variedades bravas foram selecionadas (PERRUT-LIMA; MÜHLEN; CARVALHO, 2014). Perrut-Lima e colaboradores (2014) avaliaram os níveis de toxidez de *M. esculenta* ssp. *flabellifolia* provenientes do estado de Rondônia, incluindo amostras do município de Rolim de Moura,

lugar do qual provêm a maioria dos indivíduos silvestres avaliados neste estudo. Foram constatados de moderados a elevados níveis de toxidez da raiz, permitindo que as amostras silvestres avaliadas sejam classificadas como variedades bravas. As autoras concluem que caso as amostras avaliadas sejam semelhantes às populações que deram origem à mandioca, pode-se inferir que a domesticação do cultivo ocorreu a partir de mandiocas silvestres bravas. Este cenário descarta as hipóteses (i) e (iii).

Diferentemente dos resultados encontrados por Mühlen et al. (2013), no presente estudo não foi encontrada nenhuma evidência que esteja em conformidade com a hipótese (v) de Arroyo-Kalin (2010), sendo os resultados mais congruentes com as hipóteses (ii) e (iv). Apenas no dendrograma construído para os indivíduos com base em ncSSR as variedades bravas encontram-se mais internamente em relação às variedades mansas (Figura 19). Apesar disso, antes do grande grupo de variedades mansas, e mais proximamente relacionados aos indivíduos silvestres, existem dois pequenos agrupamentos formados por um total de cinco indivíduos, dos quais apenas um é uma variedade mansa. Esta mistura de variedades nos ramos mais próximos do agrupamento de indivíduos silvestres corroboram os dendrogramas construídos com ncSSR e SNPs para as variedades mansas e bravas provenientes dos diferentes rios (Figuras 20 e 27). Na figura 20A, a princípio, chama atenção que as variedades bravas da região do rio Guaporé são as menos divergentes com relação a *M. esculenta* ssp. *flabellifolia*. Embora esta região seja geograficamente mais próxima ao centro de origem da mandioca, e de onde os indivíduos silvestres foram amostrados, provavelmente o pequeno número de indivíduos (N = 2) não é representativo da região. Quando são consideradas apenas as variedades provenientes dos principais rios, o padrão de relacionamentos se torna semelhante ao obtidos com marcadores SNPs. Apesar disso, a hipótese (v) não pode ser descartada, já que podem existir variedades mansas provenientes de outras regiões ao redor do centro de domesticação que sejam geneticamente mais relacionadas à subespécie silvestre do que as variedades bravas avaliadas neste estudo.

Variedades de mandioca mansa mais proximamente relacionadas aos indivíduos silvestres também foram observadas nos dendrogramas de indivíduos construídos com base em marcadores SNPs (Figura 26). Em ambos dendrogramas duas variedades mansas do município de Oriximiná no Pará estão situadas mais externamente ao grande grupo de variedades bravas, e distantemente relacionadas às outras três variedades mansas do mesmo município. Curiosamente, estas variedades foram as únicas a apresentarem o haplótipo F (Figuras 11 e 12), o mais divergente em relação ao haplótipo B (o mais comum entre indivíduos silvestres) e

também em relação aos haplótipos C e D (mais comuns entre as variedades cultivadas). No dendrograma de indivíduos construído com marcadores ncSSR estas variedades estão mais relacionadas a variedades bravas e localizam-se no início do grande grupo onde estão misturadas variedades mansas e bravas. Muito provavelmente estes indivíduos são resultantes de eventos de hibridização entre mandiocas mansas e bravas. Ao avaliarem estes indivíduos os agricultores podem ter associado suas características a algum tipo de variedade mansa. É possível ainda que estas variedades tenham sido originadas de eventos de incorporação de plantas resultantes do cruzamento de alguma variedade cultivada com *M. esculenta* ssp. *flabellifolia*, ou algum outro parente silvestre. Isto justificaria o posicionamento mais próximo aos indivíduos silvestres nos dendrogramas construídos com marcadores SNPs.

A divergência genética e genômica encontrada entre as mandiocas mansas e bravas ($F_{ST}(\text{ncSSR}) = 0,08$; $F_{ST}(\text{SNPs}) = 0,10$) corroboram a distinção realizada por agricultores tradicionais e estão de acordo com estudos anteriores (BRADBURY et al., 2013; ELIAS et al., 2004). Apesar desta divergência, ainda é possível detectar um apreciável grau de mistura entre as variedades mansas e bravas (Figuras 17, 19, 28 e 29). A sobreposição de grupos de variedades mansas e bravas sugerem que o fluxo gênico e incorporação de plantas oriundas de reprodução sexual (DUPUTIÉ et al., 2009; MARTINS, 2001) ocorrem com certa frequência em comunidades de agricultores tradicionais ao longo dos principais rios da bacia amazônica.

Marcadores cpSSR foram eficazes na detecção de divergências históricas entre variedades cultivadas em regiões geograficamente distintas em cultivos tais como a videira (*Vitis vinifera* ssp. *sativa*, IMAZIO et al., 2006), batata-doce (*Ipomoea batatas*, ROULLIER et al., 2011), e feijão comum (*Phaseolus vulgaris*, DESIDERIO et al., 2013). Entretanto, não foi possível detectar padrões filogeográficos a partir das análises de relacionamentos de linhagens matrilineares de variedades de mandioca. Não houveram associações entre grupos de haplótipos e os principais rios da bacia amazônica (Figuras 11B e 12). Em comparação com os trabalhos supracitados, fatores que podem ter contribuído para a ausência de padrões filogeográficos são a menor escala geográfica avaliada no presente estudo, e a origem do cultivo a partir de um evento único de domesticação. Apesar disso, como demonstrado na Figura 12 ocorre o compartilhamento de haplótipos entre as variedades de mandioca do alto rio Negro e as variedades do oeste do Pará. Estes resultados podem ser explicados pela ocorrência de mutações recorrentes (homoplasia), ou pela introdução de variedades de uma região para a outra.

A ausência de estrutura haplotípica entre as variedades provenientes de diferentes rios é refletida somente até certo ponto nas análises realizadas com os marcadores ncSSR e SNPs. As

baixas estimativas de divergência ($F_{ST}(\text{ncSSR}) = 0,025$) e SNPs ($F_{ST}(\text{SNPs}) = 0,013$), e a ausência de agrupamentos de indivíduos correspondentes aos rios nos dendrogramas das Figuras 19 e 26 corroboram o compartilhamento da maioria das linhagens haplotípicas entre os rios. De fato, divergências genéticas elevadas entre variedades de diferentes rios são observadas somente quando se comparam as variedades mansas de um rio com as variedades bravas de outro, e vice-versa. Isto é ilustrado pelas estimativas de F_{ST} par-a-par (Figuras 18 e 23), e pela maior proporção de variação observada nas AMOVAs entre variedades mansas e bravas dentro de rios (Tabelas 11 e 13) em relação à variação observada entre rios. Entretanto, quando a estruturação genética é avaliada dentro de variedades mansas e bravas separadamente, por meio de análises Bayesianas e multivariadas, alguns padrões e tendências podem ser observados (Figuras 16, 31A e 32A). Para as variedades bravas um agrupamento geográfico de variedades consistente é observado na região do alto e médio rio Negro, para marcadores ncSSR e SNPs (incluindo aqueles com comportamento *outlier*). Nestas análises as mandiocas bravas do alto e médio rio Negro possuem relacionamentos variáveis com mandiocas bravas do rio Madeira e do rio Amazonas. O mesmo padrão de estruturação genética foi encontrado para as variedades de mandioca brava provenientes do médio e alto rio Negro por Mühlen et al. (2013). Por outro lado, não foram observadas associações marcantes entre a estrutura genética de variedades mansas e as regiões geográficas. Em comum entre os marcadores ncSSR e SNPs neutros há a tendência de que variedades mansas do médio rio Negro formem um grupo mais relacionado entre si. Enquanto para os marcadores ncSSR observa-se a tendência de agrupamento entre variedades mansas dos rios Solimões e Amazonas, para os marcadores SNPs as variedades do médio e baixo rio Amazonas tendem a ser alocadas em agrupamentos diferentes. As variedades de mandioca do rio Negro podem ser distintas por serem adaptadas aos solos arenosos, altamente ácidos e inférteis da região. Entretanto, isto não parece ser regra geral, visto que as variedades mansas do alto rio Negro aparentemente não são proximamente relacionadas às variedades do médio rio Negro.

É possível que os padrões de estruturação da diversidade genética e genômica ao longo dos principais rios da bacia Amazônica estejam, pelo menos em parte, associados à diversidade de grupos étnicos pré-Colombianos que estavam estabelecidos ao longo destes rios. Dados históricos e etnográficos demonstram a importância da mandioca para os grupos indígenas amazônicos (ISENDAHL, 2011). Heckenberger (1998), por exemplo, considera que o cultivo, distribuição e consumo de mandioca foram fatores essenciais para a evolução de sociedades amazônicas complexas. Apesar de que a expansão dos grupos indígenas amazônicos deve ter

ocorrido muito após a dispersão da mandioca (ERIKSEN, 2011), é possível que diferentes grupos étnicos tenham desenvolvido conjuntos de variedades de mandioca distintos em decorrência de diferentes preferências culturais (HECKLER; ZENT, 2008; PEÑA-VENEGAS et al., 2014; RIVAL; McKEY, 2008). Desta forma, poderiam ser criados grupos de variedades divergentes entre regiões, resultantes de pressões seletivas distintas (LADIZINSKY, 1998). Eriksen (2011) sumariza uma grande quantidade de informações históricas, arqueológicas e ecológicas sobre a grande diversidade etnolinguística na Amazônia pré-Colombiana. Considerando especificamente as regiões avaliadas neste estudo, Eriksen (2011) ilustra a ocorrência dos grupos etnolinguísticos ao longo dos principais rios amazônicos. Os povos Arawak prevalentes ao longo do rio Negro, em conjunto com as etnias Yanomami, Tukano e Makú, podem ter sido os responsáveis pela diferenciação das variedades bravas do alto e médio rio Negro e pela tendência de maior relacionamento entre as variedades mansas do médio rio Negro. A presença de Arawaks também no alto rio Amazonas pode estar associada com o maior relacionamento das variedades bravas desta região com as variedades do alto e médio rio Negro na análise com ncSSRs (Figura 16B). Mas a ausência destes povos no médio rio Madeira, ocupado com maior prevalência por Tupis e Pano, não explica o maior relacionamento das variedades desta região identificada na análise de SNPs *outliers* (Figura 31B). A tendência de variedades bravas do rio Madeira serem mais proximamente relacionadas às variedades bravas do rio Amazonas (Figuras 16B e 32A) pode estar associada a ocorrência pré-Colombiana de povos Tupis ao longo destes rios. A predominância de povos Tupis ao longo do rio Solimões também pode estar relacionada à tendência de maior similaridade entre variedades mansas desta região e do rio Amazonas observada na análise com ncSSRs (Figura 16C).

5.3 Marcadores SNPs hipoteticamente sob seleção

Os estudos com marcadores moleculares sempre tiveram um impacto muito grande nos campos da ecologia, evolução e conservação (MORIN et al., 2004). Mas o estabelecimento das técnicas de sequenciamento de nova geração iniciou uma revolução ao permitir a avaliação da variação genômica de populações com base em milhares de marcadores SNPs (FUNK et al., 2012). Dentre as maiores vantagens das abordagens genômicas estão o aumento da precisão das estimativas de parâmetros genéticos, e a possibilidade de detectar locos que apresentem comportamento diferenciado (*outliers*) com relação à variação observada (LUIKART et al., 2003). Uma das metodologias comumente aplicadas para identificação de locos *outliers*, e que

foi utilizada neste estudo, baseia-se na relação entre valores de F_{ST} e heterozigosidade esperada dos locos. As estimativas de F_{ST} podem ser afetadas por todos os fatores evolutivos, dos quais seleção e mutação tendem a ser idiossincráticos a cada loco do genoma (WHITLOCK, 2011). Entretanto, a maioria dos locos autossômicos é influenciada de uma maneira aproximadamente uniforme, em maior parte pela ação da deriva (WHITLOCK, 2011). Visto que as estimativas de F_{ST} possuem pressupostos do modelo neutro (WHITLOCK; McCAULEY, 1999), e que dados genômicos permitem estimativas para um grande número de locos, existe a possibilidade de se construir uma distribuição neutra para valores de F_{ST} . A partir desta distribuição pôde-se identificar os locos *outliers* que apresentaram desvios significativos do comportamento esperado, e que possivelmente estão sob influência de seleção (BEAMOUNT; BALDING, 2004).

A domesticação de plantas oferece excelentes modelos para o estudo de processos evolutivos devido, dentre outros fatores, ao papel fundamental que a seleção teve na origem dos cultivos (MEYER; PURUGGANAN, 2013; PURUGGANAN; FULLER, 2009). Desta forma, a detecção de locos *outliers*, hipoteticamente sob seleção, nas abordagens genômicas em plantas domesticadas pode estar associada com a identificação de regiões genômicas associadas com o processo inicial de origem da domesticação e subsequente diversificação dos cultivos. Os 658 locos SNPs *outliers* que foram identificados podem, portanto, estar associados às regiões do genoma que apresentem assinaturas da seleção realizada em variedades cultivadas de mandioca a partir da subespécie silvestre que deu origem ao cultivo. Além disso, ainda foram detectados 202 locos SNPs que podem estar especificamente associados a regiões genômicas que foram alvo da seleção divergente para os grupos de variedades de mandioca mansas e bravas (McKEY et al., 2010b). Isto pode ser ilustrado pelas análises de agrupamento dos indivíduos e estruturação genômica realizadas com estes marcadores. O dendrograma construído com base nos SNPs *outliers* apresenta os indivíduos silvestres menos divergentes entre si, e mais divergentes com relação as variedades cultivadas do que o dendrograma construído com SNPs neutros (Figura 26). As DAPCs realizadas com base em marcadores SNPs *outliers* resultaram em melhor distinção entre as variedades mansas e bravas, e entre estas e os indivíduos silvestres, do que as análises realizadas com base nos marcadores SNPs neutros (Figura 30). Estas diferenças podem ser relacionadas com o método de identificação de locos *outliers* utilizados neste estudo, que é baseada na relação entre valores de F_{ST} e heterozigosidade dos locos (BEAUMONT; NICHOLS, 1996). Os locos *outliers* que apresentam desvios significativos nas estimativas de F_{ST} maximizam as divergências entre grupos, tornando a

estruturação genética mais evidente.

A distribuição geográfica da variação genômica para SNPs *outliers* (Figura 25) indica que as variedades bravas possuem maior variação alélica hipoteticamente sobre seleção ao longo dos rios do que as variedades mansas. Este resultado pode ser reflexo da predominância das variedades bravas do que variedades mansas na porção central da bacia amazônica (McKey; BECKERMAN, 1993). Maior concentração de alelos é observada na região do alto e médio rio Negro para variedades bravas, corroborando os resultados observados para os marcadores ncSSR neste estudo e também em Mühlen et al. (2013), e o mapa de ocorrência e diversidade de variedades de mandioca da Figura 4. Para as variedades mansas, maior concentração de alelos é observada na região do rio Branco. Ambos os grupos de variedades possuem áreas de maior concentração ao longo do rio Amazonas, podendo estar relacionadas ao maior período de seleção para o cultivo, visto que esta região contém as evidências arqueológicas mais antigas de ocupação humana na Amazônia (ERIKSEN, 2011). Entretanto, o registro arqueológico sugere agricultura baseada em mandioca nesta região apenas a cerca de 3.600 anos antes do presente (ERIKSEN, 2011; OLIVER, 2008), e análises paleobiológicas não contam com registros para o termo “mandioca” nesta região (BROWN et al., 2013).

As análises de estruturação genômica realizadas por meio de DAPCs permitiram que fossem identificados 97 marcadores SNPs responsáveis pela maior parte da variação observada entre os grupos de variedades avaliadas. As posições destes marcadores foram localizadas no genoma parcial da mandioca *Manihot esculenta* v4.1, que apesar de incompleto possui 30.666 locos contendo transcritos codificantes de proteínas ao longo dos seus 12.977 *scaffolds* (PROCHNIK et al., 2012). A partir deste genoma parcial foi verificado que 57 dos 97 marcadores SNPs estavam em regiões transcritas do genoma, sendo que para 16 SNPs neutros e 24 SNPs *outliers* haviam anotações funcionais definidas por termos GO. A anotação GO utiliza palavras-chave que sumarizam o conhecimento sobre as atividades funcionais de proteínas, os processos ou funções biológicas mais gerais nos quais as atividades específicas das proteínas estão inseridas, e a localização celular onde os processos ocorrem (BLAKE, 2013). Desta forma, os termos GO foram utilizados neste estudo para avaliar as funções das proteínas codificadas pelos genes nos quais os SNPs estavam inseridos.

De modo geral, foram encontradas anotações funcionais diversas para os produtos dos genes nos quais os SNPs estavam inseridos. Mesmo estando os SNPs neutros hipoteticamente sob maior influência de deriva, enquanto SNPs *outliers* estão hipoteticamente sob seleção, houveram algumas coincidências nas anotações associadas aos transcritos (embora os termos

GO nem sempre tenham sido iguais). Além disso, também houveram coincidências nas anotações quando se consideram a divergência entre as variedades cultivadas e silvestres e a distinção entre variedades mansas e bravas. Anotações envolvendo o metabolismo e transporte de moléculas foram encontradas para transcritos nos SNPs neutros relacionados tanto à divergência entre variedades cultivadas e indivíduos silvestres, como à divergência entre variedades mansas e bravas. Já os marcadores SNPs *outliers* estavam localizados em transcritos que apresentaram funções menos coincidentes. Entretanto, transcritos associados a processos catalíticos foram encontrados tanto para os marcadores responsáveis pela divergência entre variedades cultivadas e indivíduos silvestres, como para a divergência entre variedades mansas e bravas.

Recentemente, estudos de mapeamento de QTLs (locos de caráter quantitativo), estudos de associação do genoma inteiro (GWAS), e sequenciamento do genoma inteiro contribuíram para a identificação de genes associados ao início do processo de domesticação e diversificação de cultivos (MEYER; PURUGANNAN, 2013). De fato, vários genes alvo do processo de domesticação já foram identificados e caracterizados, mas grande parte deles se relacionam à origem do cultivo de grãos e leguminosas (DOEBLEY; GAUT; SMITH, 2006). Estes genes influenciam, dentre outras características, o tamanho, forma e sabor dos frutos, a cor, dormência e dispersão de sementes, e a ramificação e modo de crescimento dos caules (MEYER; PURUGGANAN, 2013; TANG; SEZEN; PATERSON, 2010). Além disso, os genes alvo do processo de domesticação codificam produtos com funções diversas, mas principalmente relacionadas a regulação da transcrição, sinalização celular, funções enzimáticas e transportadores de proteínas (DOEBLEY; GAUT; SMITH, 2006; GEPTS, 2004; MEYER; PURUGGANAN, 2013).

Dentre os locos avaliados neste estudo, L192 (SNP neutro que explica 0,41% da variação observada entre variedades mansas e bravas), e L1685 (SNP *outlier* que explica 3,16% da variação observada entre agrupamentos de variedades bravas) estão localizados em genes com termos GO associados à regulação da transcrição. Por sua vez, os locos SNPs neutros L390 e L392 (que explicam 0,47% e 0,28% da variação observada entre variedades mansas e bravas) e o loco SNP *outlier* L1406 (que explica 1,93% da variação entre variedades cultivadas e silvestres) estavam localizados em regiões gênicas não traduzidas 5'-UTR. Além destes, os locos SNPs outliers L647 e L4729 (que explicam 1,75% e 2,04% da variação observada entre variedades cultivadas e indivíduos silvestres) estavam localizados em regiões não traduzidas 3'-UTR. Regiões transcritas no RNA mensageiro (mRNA), mas não traduzidas (5'-UTR e 3'-

UTR) possuem papel importante na regulação pós-transcricional da expressão gênica (MIGNONE et al., 2002). Estas regiões contêm diferentes características estruturais, tais como alças (*stem-loops*) e regiões complementares à proteínas ligadoras de RNA, que atuam no controle da tradução, degradação e localização do mRNA (MIGNONE et al., 2012). Meyer e Purugganan (2013) observaram que mutações em elementos *cis*-reguladores (regiões, em geral não-codificantes e ligadas a um gene, que controlam os níveis ou padrões de expressão do gene) são a segunda maior classe de SNPs (após as mutações sem sentido) em genes alvo do processo de domesticação. Mutações em elementos *cis*-reguladores foram importantes, por exemplo, para a domesticação do sorgo (*Sorghum bicolor*) e do milho (*Zea mays* ssp. *mays*). No sorgo uma deleção no elemento *cis*-regulador do gene SH1 foi o responsável pela retenção dos grãos nas inflorescências (LIN et al., 2012), enquanto no milho houve a inserção do elemento transponível *Hopscotch* que causou uma mutação *cis*-reguladora no gene *tb1* (*teosinte branched 1*), alterando a arquitetura do caule (STUDER et al., 2011).

Nenhum dos locos SNPs avaliados estavam localizados em genes associados diretamente à sinalização celular. Entretanto, os locos SNPs *outliers* L4405 (que explica 3,76% da variação observada entre variedades mansas e bravas) e L1661 (que explica 2,79% da variação observada entre grupos de variedades bravas) estão localizados em genes associados ao processo de fosforilação de proteínas, com domínios funcionais envolvidos com atividade de proteína-quinase. Embora estejam associadas com uma grande variedade de processos celulares, as proteína-quinases comumente estão envolvidas também com processo de sinalização celular (NELSON; COX, 2011). Alguns locos, tais como os SNPs neutros L344, L4404 e L4576, e outros nove SNPs *outliers*, estavam localizados em transcritos associados à atividades catalíticas diversas. Apenas dois locos SNPs, L4833 (neutro) e L2210 (*outlier*), estavam em regiões gênicas diretamente associadas ao transporte de proteínas.

Os genes identificados neste estudo nos quais estão localizados marcadores SNPs neutros podem estar associados a regiões do genoma que estão sendo influenciadas pela deriva genética, contribuindo, por exemplo, para os padrões de estruturação populacional. Por outro lado, devido à uma série de fatores é difícil fazer uma relação direta entre os genes em que se localizam os SNPs *outliers* e as síndromes de domesticação da mandioca ou a divergência entre variedades mansas e bravas. Primeiramente, o conjunto de genes do genoma parcial da mandioca foi determinado por métodos de predição gênica (PROCHNIK et al., 2012), e as anotações funcionais se baseiam na similaridade dos genes preditos com genes de outras espécies. Além disso, os genes nos quais os SNPs *outliers* estavam inseridos podem não estar

diretamente relacionados com um fenótipo selecionado, mas em desequilíbrio de ligação com genes relacionados às características diretamente selecionadas (MEYER; PURUGGANAN, 2013). Outro fator importante é que a natureza da domesticação de plantas pelos humanos envolve um amplo espectro de mudanças evolutivas nas populações domesticadas, indo desde os estágios iniciais da seleção até o processo de difusão do cultivo (MEYER; PURUGGANAN, 2013). Nesse sentido, pode ser difícil diferenciar as características que foram selecionadas durante o período inicial de domesticação daquelas que surgiram durante o período de diversificação do cultivo. Estas características são variáveis entre as espécies e dependem da natureza e do número de eventos de domesticação (MEYER; PURUGGANAN, 2013; PURUGGANAN; FULLER, 2009). Enquanto as características importantes para os estágios iniciais devam envolver maior facilidade do cultivo e melhor adaptação aos ambientes antropizados, as características resultantes do processo de diversificação estão relacionadas com seleção para usos e preferências específicas (MEYER; PURUGGANAN, 2013).

A identificação dos genes da domesticação é essencial para estudos de genética de populações relacionados com o processo de domesticação e com o entendimento dos fatores evolutivos que vêm afetando os cultivos, tais como *selective sweeps* e fluxo gênico (GEPTS, 2004). As informações sobre variação funcional geradas neste estudo podem ser de grande valia para avaliações mais aprofundadas sobre os processos evolutivos que vêm ocorrendo com as populações cultivadas de mandioca desde sua domesticação. O conjunto de marcadores SNPs com comportamento *outlier* detectados neste estudo podem ser utilizados como ponto de partida para a determinação dos genes mais importantes para a domesticação da mandioca, divergência entre variedades mansas e bravas, e também em questões de adaptação local, como por exemplo cultivo em diferentes ambientes.

5.4 Considerações sobre a dispersão do cultivo

A ampla distribuição do haplótipo cpSSR mais comum entre variedades mansas e bravas corrobora uma ampla dispersão da mandioca ao longo dos rios amazônicos muito antes da chegada dos europeus no continente americano, assim como foi observado em batata-doce (ROULLIER et al., 2011). A mandioca deve ter sido ampla e rapidamente dispersa a partir de sua domesticação, que pode ter ocorrido até 10 mil anos antes do presente (AP) (OLSEN; SCHAAL, 1999). Apesar da relativa escassez de informações (PICKERSGILL, 2007), o registro arqueológico e paleobotânico sugere uma ampla dispersão da mandioca na maior parte dos

Neotrópicos já há 6.500 AP (ISENDAHL, 2011). Esta ampla dispersão é justificada considerando que os vestígios mais antigos de mandioca cultivada de que se tem conhecimento foram encontrados na costa pacífica do Peru, e sugerem que a mandioca cruzou os Andes há cerca de 8.000 AP (ISENDAHL, 2011). Análises das proto-linguagens para quais é possível reconstruir o termo “mandioca” sugerem uma ampla distribuição na América do Sul e Mesoamérica antes mesmo do período de estabelecimento de sociedades baseadas na agricultura, entre 4.000 e 3.000 AP (BROWN et al., 2013). Estimativas da antiguidade do cultivo de mandioca ao longo dos rios avaliados neste estudo podem ser retiradas da revisão de várias linhas de evidência feita por Eriksen (2011). Devido à proximidade com o centro de domesticação, a mandioca deve ter sido dispersa ao longo do rio Madeira tão cedo quanto 9.000 AP. Enquanto Eriksen (2011) não comenta sobre as regiões dos rios Negro e Branco, o cultivo de mandioca pode ter sido amplamente disperso ao longo do rio Amazonas há cerca de 5.700 AP. Na porção ocidental da Amazônia, incluindo o rio Solimões, existem evidências do cultivo de mandioca a 4 mil anos AP (ERIKSEN, 2011).

Os padrões contemporâneos de variação e estrutura genética de plantas cultivadas refletem processos históricos associados com a domesticação, tais como a origem geográfica dos cultivos e os seus padrões de dispersão e difusão (MILLER; SCHAAL, 2005; ROULLIER et al., 2011). No caso dos cultivos Amazônicos é provável que os padrões de dispersão tenham sido determinados, dentre outros fatores, por preferências étnicas (CLEMENT et al., 2010). Neste estudo padrões mais claros de estrutura genética e genômica foram observados entre variedades mansas e bravas do que entre rios. No entanto, quando a estruturação é avaliada separadamente para cada grupo de variedades alguns padrões e tendências foram observados. Como discutido no tópico 5.2 é possível que diferentes preferências étnicas estejam associadas à diferenciação das variedades bravas do alto e médio rio Negro e pelos relacionamentos variáveis entre mandiocas mansas.

Por outro lado, é possível também que as assinaturas de diferentes dispersões tenham sido, pelo menos em parte, obscurecidas por processos históricos mais recentes, como observado em variedades de batata-doce da Oceania (ROULLIER et al., 2013). Os rios avaliados neste estudo tiveram grande importância como centros e regiões de diversidade de cultivos amazônicos no período pré-Colombiano. Estas áreas estavam relacionadas com altas densidades de populações humanas e intensificação da agricultura (CLEMENT, 1999a), constituindo importantes centros de diversidade genética de muitos cultivos (CLEMENT et al., 2010). Através da troca de variedades de mandiocas, os centros e regiões de diversidade

poderiam ter colaborado para a homogeneização dos conjuntos de variedades desenvolvidos por diferentes grupos étnicos pré-Colombianos. Outro fator que pode ter colaborado para a falta de padrões mais claros de estruturação genética de variedades de mandiocas mansas e bravas entre rios foi a interferência que a conquista europeia teve sobre as populações indígenas amazônicas. Estima-se que entre 90 e 95% da população indígena tenha sido dizimada pela introdução de doenças, guerras e escravidão europeia nos dois séculos seguintes à colonização (ERIKSEN, 2011). Esta drástica redução populacional pode estar relacionada à uma perda apreciável da diversidade de cultivos, e provocou o deslocamento e fuga de populações indígenas sobreviventes (CLEMENT, 1999b). O movimento destes grupos indígenas pode ter colaborado para a alteração dos padrões de distribuição da diversidade genética de variedades de mandioca ao longo dos rios amazônicos. Ainda mais recentemente, o ciclo da borracha causou uma migração em massa da região nordeste do Brasil para diferentes partes da Amazônia. Possivelmente os migrantes contribuíram para a difusão de variedades de mandioca menos relacionadas às variedades amazônicas, o que também pode ter colaborado para a redução do sinal de estruturação genética das variedades de mandioca entre os rios avaliados neste estudo. Nesse sentido, abordagens como a de Roullier e colaboradores (2013), que usaram amostras de DNA antigo depositados em herbários, poderia ser útil na elucidação da possível atuação que eventos mais recentes tiveram nos padrões contemporâneos de distribuição da variação e estrutura genética de variedades de mandioca.

A ausência de padrões claros de estruturação genética entre os rios observado neste estudo não permite que sejam feitas inferências diretas sobre como se deu o processo de dispersão dos grupos de variedades de mandioca. Ainda assim, o simples fato de que os padrões de estruturação genética de variedades mansas e bravas tendem a ser distintos pode estar relacionado a histórias de dispersão distintas. Mesmo que não tenham sido observados resultados que corroborem a hipótese de Arroyo-Kalin (2010), ela não pode ser descartada com base neste estudo. Caso as variedades mansas tenham sido as primeiras a serem criadas, pode ser que quando as variedades bravas foram desenvolvidas as variedades mansas já estivessem bem dispersas. Neste cenário as variedades mansas teriam maior tempo de diversificação, justificando a maior diversidade alélica encontrada neste trabalho para este grupo. Considerando que as variedades bravas teriam sido desenvolvidas com a intensificação da agricultura na Amazônia entre 4.000 e 3.000 AP, as mandiocas bravas tiveram importância fundamental para o sustento e para o desenvolvimento de sociedades pré-Colombianas na Amazônia (HECKENBERGER, 1998). A partir deste ponto as variedades bravas teriam se

tornado a principal cultura básica dos povos Amazônicos, justificando a maior ocorrência nesta região em relação às mandiocas mansas (McKEY; BECKERMAN, 1993). O impacto de pressões seletivas e fatores evolutivos seriam diferentes entre os grupos de variedades nas áreas em que as variedades mansas são cultivos secundários em relação às variedades bravas, como é o caso da maior parte da Amazônia brasileira. Com isso as mandiocas bravas estariam mais sujeitas à seleção divergente entre regiões resultante de preferências culturais distintas, enquanto as mandiocas mansas teriam divergências mais sutis entre regiões. Dessa forma, os processos que ocorreram durante a história do cultivo de variedades bravas resultariam em padrões de estrutura genética contemporâneos mais evidentes do que as variedades mansas.

Este estudo foi um esforço no sentido de se relacionar os padrões de distribuição da variação e estruturação genética e genômica de variedades de mandioca cultivadas tradicionalmente na Amazônia com a história de dispersão do cultivo. Investigando diferentes marcadores moleculares com caráter histórico (cpSSRs), contemporâneo (ncSSR) e genômico (SNPs) novas informações puderam ser geradas e outras questões podem ser levantadas. (i) Poderiam estudos que avaliassem variedades de mandioca cultivadas em outras regiões e rios ser úteis na detecção de estrutura genética regional? Nesse sentido pode ser útil a avaliação da estrutura genética e dos relacionamentos entre linhagens matrilineares em variedades de mandioca provenientes de outras regiões amazônicas, com base em marcadores ncSSR, cpSSR e sequenciamento de genes cloroplastidiais. (ii) Até que ponto os genes nos quais estão localizados os marcadores SNPs *outliers* avaliados contribuíram para domesticação da mandioca e a divergência entre variedades mansas e bravas? Quais seriam as características que estes genes regulam? Para tanto, podem ser realizadas abordagens de mapeamento genético, genética de associação e genética reversa visando a determinação dos locais e funções gênicas associados a características relacionadas as síndromes de domesticação e diversificação do cultivo. (iii) Qual o nível de variação funcional existente nos genes associados à divergência entre variedades mansas e bravas de diferentes regiões? O desenho de iniciadores específicos para os genes nos quais os SNPs identificados neste estudo estão inseridos, permitiria que estas regiões pudessem ser amplificadas e avaliadas quanto ao seu nível de variação molecular, com base na ocorrência de marcadores SNPs dentro das sequências destes genes. Para a avaliação simultânea de um maior número de genes, e maior precisão estatística, abordagens de sequenciamento de nova geração poderiam ser utilizadas para a aquisição de grande quantidade de informação.

6 CONCLUSÃO

Com base nos resultados obtidos e nos objetivos e hipóteses apresentadas neste estudo, pode-se concluir que:

- 1 - As comunidades de agricultores tradicionais ao longo dos principais rios da bacia amazônica manejam elevados níveis de diversidade genética e genômica em suas variedades de mandioca mansas e bravas;
- 2 – Caso a diversidade genética de indivíduos silvestres seja representativa para a espécie *Manihot esculenta* ssp. *flabellifolia*, não há indícios de deriva genética associada à domesticação da mandioca, possivelmente em função das práticas do manejo tradicional de variedades de mandioca cultivada na região amazônica;
- 3 - Não houve associações de linhagens matrilineares com os grupos de variedades de mandioca mansas e bravas, nem com os diferentes rios da bacia amazônica, muito provavelmente resultante da ancestralidade comum dos grupos de variedades cultivadas com a espécie *Manihot esculenta* ssp. *flabellifolia*;
- 4 - Existe uma consistente divergência genética entre variedades de mandioca mansas e bravas. Por outro lado, não há indícios de estruturação genética aparente entre grupos de variedades provenientes dos principais rios. No entanto, os padrões de estrutura genética dentro de variedades mansas e bravas tendem a ser contrastantes;
- 5 - Os perfis de variação genômica indicam que não há divergência marcante entre variedades provenientes de diferentes rios, sendo a divergência entre variedades de mandioca mansas e bravas a mais marcante. Existem locos associados às regiões do genoma que hipoteticamente estão sob seleção antrópica tanto para a domesticação do cultivo a partir do seu ancestral silvestre, como para a seleção divergente entre as mandiocas mansas e bravas;
- 6 - Devido à ausência de padrões claros de estrutura genética e genômica consistente entre variedades provenientes de diferentes rios não é possível fazer inferências sobre prováveis rotas de dispersão e difusão do cultivo a partir do seu centro de origem no sudoeste da bacia amazônica brasileira. No entanto, os padrões contrastantes de diferenciação genética e genômica dentro de variedades mansas e bravas podem estar associados a histórias de dispersão distintas para as variedades mandioca.

REFERÊNCIAS

- ALLEM, A.C. The origin of *Manihot esculenta* Crantz (Euphorbiaceae). **Genetic Resources and Crop Evolution**, Dordrecht, v. 41, p. 133-150, 1994.
- ALLEM, A.C. The origins and taxonomy of cassava. In: HILLOCKS, R.J.; THRESH, J.M.; BELLOTTI, A.C. (Ed). **Cassava: biology, production and utilization**. Oxford: CAB International, 2002. p. 1-16.
- ALLEM, A.C.; MENDES, R.A.; SALOMÃO, A.N.; BURLE, M.L. The primary gene pool of cassava (*Manihot esculenta* Crantz subspecies *esculenta*, Euphorbiaceae). **Euphytica**, Wageningen, v. 120, p. 127-132, 2001.
- ALLENDORF, F.W.; HOHENLOHE, P.A.; LUIKART, G.H. Genomics and the future of conservation genetics. **Nature Reviews Genetics**, London, v. 11, p. 697-709, 2010.
- ALLENDORF, F.W.; LUIKART, G.H.; AITKEN, S.N. **Conservation and the genetics of populations**. 2nd ed. Oxford, Wiley-Blackwell, 2013. 624 p.
- ALVES, A.A.C. Cassava Botany and Physiology. In: HILLOCKS, R.J.; THRESH, J.M.; BELLOTTI, A.C. (Ed). **Cassava: biology, production and utilization**. Oxford: CAB International, 2002. p. 67-89.
- ALVES-PEREIRA, A.; PERONI, N.; ABREU, A.G.; GRIBEL, R.; CLEMENT, C.R. Genetic structure of traditional varieties of bitter manioc in three soils in Central Amazonia. **Genetica**, Gravenhage, v. 139, p. 1259-1271, 2011.
- ANGIOI, S.E.; RAU, D; ATTENE, G.; NANNI, L.; BELLUCCI, E.; LOGOZZO, G.; NEGRI, V.; SPAGNOLETTI ZEULI, P.L.; PAPA, R. Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. **Theoretical and Applied Genetics**, Berlin, v. 121, p. 829-843, 2010.
- ANTAO, T.; LOPES, A.; LOPES, R.J.; BEJA-PEREIRA, A.; LUIKART, G. LOSITAN: A workbench to detect molecular adaptation based on a F_{ST} -outlier method. **BMC Bioinformatics**, Oxford, v. 9, p.323, 2008.
- ARROYO-KALIN, M. Steps towards an ecology of landscape: The pedo-stratigraphy of anthropogenic dark earths. In: WOODS, W.I.; TEIXEIRA, W.G.; LEHMANN, J.; STEINER, C.; WINKLERPRINS, A.M.G.A.; REBELLATO, L. (Ed). **Amazonian Dark Earths: Wim Sombroek's Vision**. New York: Springer, 2009. p. 33-83.
- _____. The Amazonian Formative: Crop domestication and anthropogenic soils. **Diversity**, Basel, v. 2, p. 473-504, 2010.
- AVISE, J.C. **Molecular markers, natural history, and evolution**. 2nd ed. Massachussets, Sinauer Associates, 2004. 684 p.
- _____. Phylogeography: retrospect and prospect. **Journal of Biogeography**, Edinburgh, v. 36, p. 3-15, 2009.

BAIRD, N.A.; ETTER, P.D.; ATWOOD, T.S.; CURREY, M.C.; SHIVER, A.L.; LEWIS, Z.A.; SELKER, E.U.; CRESKO, W.A.; JOHNSON, E.A. Rapid SNP discovery and genetic mapping using sequenced RAD markers. **PLoS ONE**, San Francisco, v. 3, e3376, 2008.

BANDELT, H.J.; FORSTER, P.; RÖHL, A. Median-Joining Networks for Inferring Intraspecific Phylogenies. **Molecular Biology and Evolution**, Chicago, v. 16, n. 1, p. 37-48, 1999.

BARCHI, L.; LANTERI, S.; PORTIS, E.; ACQUADRO, A.; VALÈ, G.; TOPPINO, L.; ROTINO, G.L. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. **BMC Genomics**, Austin, v. 2:304, doi:10.1186/1471-2164-12-304, 2011.

BEAUMONT, M.A. Adaptation and speciation: what can F_{ST} tell us? **Trends in Ecology & Evolution**, Cambridge, v. 20, n. 8, p. 435-440, 2005.

BEAMOUNT, M.A.; BALDING, D.J. Identifying adaptive genetic divergence among populations from genome scans. **Molecular Ecology**, Oxford, v. 13, n. 4, p. 969-980, 2004.

BEAUMONT, M.A.; NICHOLS, R. A. Evaluating loci for use in the genetic analysis of population structure. **Proceedings of the Royal Society of London, Series B**, London, v. 263, p. 1619-1626, 1996.

BENTLEY, D.R. et al. Accurate whole human genome sequencing using reversible terminator chemistry. **Nature**, London, v. 456, p. 53-59, 2008.

BITOCCHI, E.; NANNI, L.; BELLUCCI, E.; ROSSI, M.; GIARDINI, A.; ZEULI, P.S.; LOGOZZO, G.; STOUGAARD, J.; MCCLEAN, P.; ATTENE, G.; PAPA, R. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequencing data. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 109, n. 14, p. e788-e796, 2012.

BLACK, I.V.W.C.; BAER, C.F.; ANTOLIN, M.F.; DUTEAU, N.M. Population genomics: genome-wide sampling of insect populations. **Annual Review of Entomology**, Palo Alto, v. 46, p. 441-469, 2001.

BLAKE, J.A. Ten quick tips for using the Gene Ontology. **PLoS Computational Biology**, San Francisco, v. 9, n. 11, e1003343, doi:10.1371/journal.pcbi.1003343, 2013.

BOSTER, J.S. Exchange of varieties and information between Aguaruna manioc cultivators. **American Anthropologist**, Arlington, v. 88, n.2, p. 428-436, 1986.

_____. Selection for perceptual distinctiveness: evidence from Aguaruna cultivar of *Manihot esculenta*. **Economic Botany**, New York, v. 39, n. 3, p. 310-325, 1985.

BOCCACCI, P.; BOTTA, R. Investigating the origin of hazelnut (*Corylus avellana* L.) cultivars using chloroplast microsatellites. **Genetic Resources and Crop Evolution**, Dordrecht, v. 56, p. 851-859, 2009.

BOKANGA, M.; EKANAYAKE, I.J.; DIXON, A.G.O.; PORTO, M.C.M. Genotype-

environment interactions for cyanogenic potential in cassava. **Acta Horticulturae**, v. 375, p. 131-139, 1994.

BONIERBALE, M.; ROA, A.C.; MAYA, M.M.; DUQUE M.C.; THOME, J. Assessment of genetic diversity in *Manihot* species with AFLPs. **African Journal of Root and Tuber Crops**, v. 2, p. 139, 1997.

BRADBURY, E.J.; DUPUTIÉ, A.; DELÊTRE, M.; ROULLIER, C.; TRUJILLO, A.N.; MANU-ADUENING, J.W.; EMSHWILLER, E.; McKEY, D. Geographic differences in patterns of genetic differentiation among bitter and sweet cassava (*Manihot esculenta*: Euphorbiaceae). **American Journal of Botany**, Baltimore, v. 100, n. 5, p.857-866, 2013.

BROWN, A.H.D.; WEIR, B.S. Measuring genetic variability in plant populations. In: TANSKLEY, S.D.; ORTON, T.J. (Ed). **Isozymes in Plant Genetics and Breeding**. Amsterdam: Elsevier, 1983. p. 219-239.

BROWN, C.H.; CLEMENT, C.R.; EPPS, P.; LUEDELING, E.; WICHMANN, S. The Paleobiolinguistics of Domesticated Manioc (*Manihot esculenta*). **Ethnobiology Letters**, Denton, v. 4, p. 61-70, 2013.

BRUMFIELD, R.T.; BEERLI, B.; NICKERSON, D.A.; EDWARDS, S.V. The utility of single nucleotide polymorphisms in inferences of population history. **Trends in Ecology and Evolution**, Cambridge, v. 18, n.5, p. 246-256, 2003.

BRUSH, S.B.; CARNEY, H.J.; HUAMÁN, Z. Dynamics of Andean potato agriculture. **Economic Botany**, New York, v. 35, n. 1, p. 70-88, 1981.

BURNS, A.E.; GLEADOW, R.M.; ZACARIAS, A.M.; CUAMBE, C.E.; MILLER, R.E.; CAVAGNARO, T.R. Variations in the chemical composition of cassava (*Manihot esculenta* Crantz) leaves and roots as affected by genotypic and environmental Variation. **Journal of Agriculture and Food Chemistry**, v. 60, p. 4946-4956, 2012.

CALLE, F.; PEREZ, J.C.; GAITÁN, W.; MORANTE, N.; CEBALLOS, H.; LLANO, G; ALVAREZ, E. Diallel inheritance of relevant traits in cassava (*Manihot esculenta* Crantz) adapted to acid-soils savannas. **Euphytica**, Wageningen, v. 144, n. 1-2, p. 177-186, 2005.

CASTRO, I.; PINTO-CARNIDE, O.; ORTIZ, J.M.; MARTÍN, J.P. Chloroplast genome diversity in Portuguese grapevine (*Vitis vinifera* L.) cultivars. **Molecular Biotechnology**, v. 54, p. 528-540, 2013.

CAVALLI-SFORZA, L.L.; EDWARDS, A.W.F. Phylogenetics analysis: Models and estimation procedures. **American Journal of Human Genetics**, Maryland, v.19, p. 233-257, 1967.

CHACÓN, J.; MADRIÑÁN, S.; DEBOUCK, D.; RODRIGUEZ, F.; TOHME, J. Phylogenetic patterns in the genus *Manihot* (Euphorbiaceae) inferred from analyses of nuclear and chloroplast DNA regions. **Molecular Phylogenetics and Evolution**, San Diego, v. 49, p. 260-267, 2008.

CHAVARRIAGA-AGUIRRE, P.; MAYA, M.M.; TOHME, J.; DUQUE, M.C.; IGLESIAS, C.;

- BONIERBALE, M.W.; KRESOVICH, S.K.; KOCHERT, G.K. Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundancy in the cassava core collection and to assess the usefulness of DNA-based markers to maintain germplasm collections. **Molecular Breeding**, Dordrecht, v. 5, p. 263-273, 1999.
- CHAVARRIAGA-AGUIRRE, P.; MAYA, M.M.; BONIERBALE, M.V.; KRESOLICH, S.; FREGENE, M.A.; TOEHNE, J.; KOCHERT, G. Microsatellites in Cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. **Theoretical and Applied Genetics**, Berlin, v. 97, n. 3, p. 493-501, 1998.
- CHEN, C.; MITCHELL, S.E.; ELSHIRE, R.J.; BUCKLER, E.S.; EL-KASSABY, Y.A. Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. **Tree Genetics & Genomes**, Berlin, v. 9, n. 6, p.1537-1544, 2013.
- CHUNG, S.-M.; STAUB, J.E. The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. **Theoretical and Applied Genetics**, Berlin, v. 107, n. 8, p. 757-767, 2003.
- CLEMENT, C.R. 1492 and the loss of Amazonian crop genetic resources. II. Crop biogeography at contact. **Economic Botany**, New York, v.53, n.2, p. 203-216, 1999a.
- CLEMENT, C.R. 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. **Economic Botany**, New York, v.53, n.2, p. 188-202, 1999b.
- CLEMENT, C.R.; CRISTO-ARAÚJO, M.; COPPENS D'EECKENBRUGGE, G.; PEREIRA, A.A.; PICANÇO-RODRIGUES, D. Origin and domestication of native Amazonian crops. **Diversity**, Basel, v. 2, p. 72-106, 2010.
- COLLINS, F.S.; BROOKS, L.D.; CHAKRAVARTI, A. A DNA polymorphism discovery resource for research on human genetic variation. **Genome Research**, New York, v. 8, p. 1229-1231, 1998.
- COOMES, O.T. Of stakes, stems and cuttings: the importance of local seed systems in traditional Amazonian societies. **The Professional Geographer**, London, v. 62, n. 3, p.323-334, 2010.
- DANIELL, H.; WURDACK, K.J.; KANAGARAJ, A.; LEE, S.-B.; SASAKI, C.; JANSEN, R.K. The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of *atpF* in Malpighiales: RNA editing and multiple losses of a group II intron. **Theoretical and Applied Genetetics**, Berlin, v.116, p. 723-737, 2008.
- DAVEY, J.W.; BLAXTER, M.L. RADSeq: next-generation population genetics. **Briefings in Functional Genomics**, Oxford, v. 9, p. 416-423, 2010.
- DAVEY, J.W.; HOHENLOHE, P.A.; ETTER, P.D.; BOONE, J.Q.; CATCHEN, J.M.; BLAXTER, M.L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics**, London, v. 12, p. 499-510, 2011.

DELÊTRE, M.; McKEY, D.B.; HODKINSON, T.R. Marriage exchanges, seed exchange, and the dynamics of manioc diversity. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 108, n. 45, p. 18249-18254, 2011.

DESIDERIO, F.; BITOCCHI, E.; BELLUCCI, E.; RAU, D.; RODRIGUEZ, M.; ATTENE, G.; PAPA, R.; NANNI, L. Chloroplast microsatellite diversity in *Phaseolus vulgaris*. **Frontiers in Plant Science**, Lausanne, v. 3, artigo 312, doi: 10.3389/fpls.2012.00312, 2013.

DIERINGER, D.; SCHLÖTTERER, C. Microsatellite analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. **Molecular Ecology Notes**, Oxford, v. 3, p. 167-169, 2003.

DOEBLEY, J.F.; GAUT, B.S.; SMITH, B.D. The molecular genetics of crop domestication. **Cell**, Cambridge, v. 127, p. 1309-1321, 2006.

DOYLE, J.J.; DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**, Rockville, v.1, p. 13-15, 1990.

DUPUTIÉ, A.; SALICK, J.; McKEY, D. Evolutionary biogeography of *Manihot* (Euphorbiaceae), a rapidly radiating Neotropical genus restricted to dry environments. **Journal of Biogeography**, Oxford, v. 38, n. 6, p. 1033-1043, 2011.

DUPUTIÉ, A.; DAVID, P.; DEBAIN, C.; McKEY, D. Natural hybridization between a clonally propagated crop, Cassava (*Manihot esculenta* Crantz) and wild relative in French Guiana. **Molecular Ecology**, Oxford, v. 16, p. 3025-3038, 2007.

DUPUTIÉ, A.; MASSOL, F.; DAVID, P.; HAXAIRE, C.; McKEY, D. Traditional Amerindian cultivators combine directional and ideotypic selection for sustainable management of Cassava genetic diversity. **Journal of Evolutionary Biology**, Oxford, v. 22, p. 1317-1325, 2009.

DYER, G.A.; GONZÁLEZ, C.; LOPERA, D.C. Informal “seed” systems and the management of gene flow in traditional agroecosystems: the case of cassava in Cauca, Colombia. **PLoS ONE**, San Francisco, v. 6, n. 2, e29067.

EARL, D.A.; VONHOLDT, B.M. Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. **Conservation Genetics Resources**, Dordrecht, v. 4, n.2, p. 359-361, 2012.

EBERT, D.; PEAKALL, R. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. **Molecular Ecology Resources**, Oxford, v. 9, p. 673-690, 2009.

ELIAS, M.; McKEY, D. The unmanaged reproductive ecology of domesticated plants in traditional agroecosystems: An example involving Cassava and a call for data. **Acta Oecologica**, Paris, v. 21, n. 3, p. 223-230, 2000.

ELIAS, M.; MÜHLEN, 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. **Economic Botany**, New York, v. 58, n.2, p. 242-256, 2004.

ELIAS, M.; PENET, L.; VINDRY, P.; McKEY, D.; PANAUD, O.; ROBERT, T. Unmanaged sexual reproduction and the dynamics of genetic diversity of a vegetatively propagated crop plant, Cassava (*Manihot esculenta* Crantz), in a traditional farming system. **Molecular Ecology**, Oxford, v. 10, p. 1895-1907, 2001.

ELIAS, M.; RIVAL, L.; McKEY, D. Perception and management of Cassava (*Manihot esculenta* Crantz) diversity among Makushi Amerindians of Guyana (South America). **Journal of Ethnobiology**, Denton, v. 20, n. 2, p. 239-265, 2000.

ELMOUSADIK, A.; PETIT, R.J. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. **Theoretical and Applied Genetics**, Berlin, v. 92, n. 7, p. 832-839, 1996.

EMANUELLI, F.; LORENZI, S.; GRZESKOWIAK, L.; CATALANO, V.; STEFANINI, M.; TROGGIO, M.; MYLES, S.; MARTINEZ-ZAPATER, J.M.; ZYPRIAN, E.; MOREIRA, F.M.; GRANDO, M.S. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. **BMC Plant Biology**, Austin, v. 13:39, doi:10.1186/1471-2229-13-39, 2013.

EMERSON, K.J.; MERZ, C.R.; CATCHEN, J.M.; HOHENLOHE, P.A.; CRESKO, W.A.; BRADSHAW, W.E.; HOLZAPFEL, C.M. Resolving postglacial phylogeography using high-throughput sequencing. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 107, p. 16196-16200, 2010.

EMPERAIRE, L. Elementos de discussão sobre a conservação da agrobiodiversidade: o exemplo da mandioca (*Manihot esculenta* Crantz) na Amazônia brasileira. In : CAPOBIANDO, J.-P. (Ed) **Biodiversidade da Amazônia brasileira, avaliação e ações prioritárias para a conservação, uso sustentável e repartição dos benefícios**. São Paulo : ISA, 2001. p. 225-234.

_____. A agrobiodiversidade em risco. **Ciência Hoje**, São Paulo, v. 32, n. 187, p. 28-33, 2002.

_____. L'agrobiodiversité en Amazonie brésilienne : ressource et patrimoine. **Journal d'Ethnobiologie**, Denton, v. 42, p. 413-426, 2005.

EMPERAIRE, L.; PERONI, N. Traditional management of agrobiodiversity in Brazil: a case study of Manioc. **Human Ecology**, New York, v. 35, n. 6, p. 761-768, 2007.

ERIKSEN, L. **Nature and culture in prehistoric Amazonia. Using G.I.S. to reconstruct ancient ethnogenetic process from archaeology, linguistics, geography, and ethnohistory**. Lund, Lund University, 2011. p.375.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Molecular Ecology**, Oxford, v. 14, p. 2611-2620, 2005.

EXCOFFIER, L.; LISCHER, H.E.L. Arlequin suit ver. 3.5: A new series of programas to perform population genetics analyses under Linux or Windows. **Molecular Ecology**

Resources, Oxford, v. 10, n. 6, p. 564-567, 2010.

FELSENSTEIN, J. **PHYLIP (Phylogeny Inference Package)**, version 3.6. Computer program distributed by the author, Department of Genome Sciences, University of Washington, Seattle. 2005.

FERGUSON, M.E.; HEARNE, S.J.; CLOSE, T.J.; WANAMAKER, S.; MOSKAL, W.A.; TOWN, C.D.; YOUNG, J. de; MARRI, P.R.; RABBI, I.Y.; VILLIERS, E.P. de. Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. **Theoretical and Applied Genetics**, Berlin, v. 124, n. 4, p.685-695, 2012.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3.ed. Brasília: Embrapa Cenargen, 1998. p. 220.

FORSTER, P.; BANDELT, H.J.; ROHL, A. **NETWORK 3.1.1.0**. Cambridge, 2000. Disponível em: <<http://www.fluxus-engineering.com>>. Acessado em 15 janeiro 2012.

FRASER, J.A. The diversity of bitter manioc (*Manihot esculenta* Crantz) cultivation in a whitewater Amazonian landscape. **Diversity**, Basel, v. 2, n. 4, p. 586-609, 2010.

FRASER, J.A.; ALVES-PEREIRA, A.; JUNQUEIRA, A.B.; PERONI, N.; CLEMENT, C.R. Convergent adaptations: bitter manioc cultivation systems in fertile Anthropogenic Dark Earths and floodplain soils in Central Amazonia. **PLoS ONE**, SanFrancisco, v. 7, n. 8, e43636. doi:10.1371/journal.pone.0043636, 2012.

FRASER, J.A.; CLEMENT, C.R. Dark Earths and manioc cultivation in Central Amazonia: A window on pre-Columbian agricultural systems? **Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas**, Belém, v. 3, n. 2, p. 175-194, 2008.

FREGENE, M.A.; SUAREZ, M.; MKUMBIRA, J.; KULEMBEKA, H.; NDEDYA, E.; KULAYA, A.; MITCHEL, S.; GUILBERG, U.; ROSLING, H.; DIXON, A.G.O.; DEAN, R.; KRESOVICH, S. Simple sequence repeat marker diversity in Cassava landraces: Genetic diversity and differentiation in a asexually propagated crop. **Theoretical and Applied Genetics**, Berlin, v. 107, p. 1083-1093, 2003.

FREELAND, J.R. **Molecular Ecology**. Hoboken: Wiley-Blackwell, 2005. p. 400.

FUNK, W.C.; McKAY, J.K.; HOHENLOHE, P.A.; ALLENDORF, F.W. Harnessing genomics for delineating conservation units. **Trends in Ecology and Evolution**, Cambridge, v. 27, n.9, p. 489-495, 2012.

GEPTS, P. Crop domestication as a long-term selection experiment. **Plant Breeding Reviews**, Berlin, v. 24, p. 1-44, 2004.

GOLDSTEIN, D.B.; POLLOCK, D.D. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. **Journal of Heredity**, Oxford, v. 88, p. 335-342, 1997.

GOUDET, J. Hierfstat, a package for R to compute and test hierarchical F-statistics. **Molecular Ecology Notes**, Oxford, v. 5, n. 1, p. 184-186, 2005.

GUICHOUX, E.; LAGACHE, L.; WAGNER, S.; CHAUMEIL, P.; LÉGER, P.; LEPAIS, O.; LEPOITTEVIN, C.; MALAUSA, T.; REVARDEL, E.; SALIN, F.; PETIT, R.J. Current trends in microsatellite genotyping. **Molecular Ecology Resources**, Oxford, v. 11, p. 591-611, 2011.

HECKENBERGER, M.J. Manioc agriculture and sedentism in Amazonia: the Upper Xingu example. **Antiquity**, Durham, v.72, p. 633-648, 1998.

HECKLER, S.; ZENT, S. Piara manioc varietal: hyperdiversity or social currency? **Human Ecology**, New York, v. 36, p. 679-697, 2008.

HOHENLOHE, P.A.; AMISH, S.J.; CATCHEN, J.M.; ALLENDORF, F.W.; LUIKART, G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. **Molecular Ecology Resources**, Oxford, v. 11, p. 117-122, 2011.

HOHENLOHE, P.A.; BASSHAM, S.; ETTER, P.D.; STIFFLER, N.; JOHNSON, E.A.; CRESKO, W.A. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. **PLoS Genetics**, San Francisco, v. 6, n. 2, p.e1000862, 2010.

IMAZIO, S.; LABRA, M.; GRASSI, F.; SCIENZA, A.; FAILLA, O. Chloroplast microsatellites to investigate the origin of grapevine. **Genetic Resources and Crop Evolution**, Dordrecht, v. 53, p. 1003-1011, 2006.

ISENDAHL, C. The domestication and early spread of manioc (*Manihot esculenta* Crantz): a brief synthesis. **Latin American Antiquity**, Washington, v. 22, n. 4, p. 452-468, 2011.

JAKOBSSON, M.; ROSENBERG, N.A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. **Bioinformatics**, Oxford v. 23, n. 14, p. 1801-1806, 2017.

JARAMILLO, G.; MORANTE, N.; PÉREZ, J.C.; CALLE, F.; CEBALLOS, H.; ARIAS, B.; BELLOTTI, A.C. Diallel analysis in cassava adapted to the midaltitude valleys environment. **Crop Science**, Madison, v. 45, n. 3, p. 1058-1063, 2005.

JENNINGS, D.L. Cassava. In: SMARTT, J.; SIMMONDS, N.W. (Ed). **Evolution of Crop Plants**. London: Longman, 1995. p. 128-132.

JOMBART, T.; AHMED, I. Adegnet 1.3-1: new tools for the analysis of genome-wide SNP data. **Bioinformatics**, Oxford, v. 27, n. 21, p. 3070-3071, doi: 10.1093/bioinformatics/btr521, 2011.

JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC Genetics**, Austin, v. 11:94, 2010.

KALDERIMIS, A.; LYNE, R.; BUTANO, D.; CONTRINO, S.; LYNE, M.; HEIMBACH, J.; HU, F.; SMITH, R.; ŠTĚPÁN, R.; SULLIVAN, J.; MICKLEM, G. InterMine: extensive web services for modern biology. **Nucleic Acids Research**, London, v. 42, W468-W472, doi: 10.1093/nar/gku301, 2014.

KAMVAR, Z.N.; TABIMA, J.F.; GRÜNWALD, N.J. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. **PeerJ**, San Francisco, v. 2: e281, DOI 10.7717/peerj.281, 2014.

LADIZINSKY, G. **Plant Evolution Under Domestication**. Dordrecht: Springer, 1998. p. 254.

LANGMEAD B.; TRAPNELL C.; POP M.; SALZBERG, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biology**, London, v. 10, R25, DOI:10.1186/gb-2009-10-3-r25, 2009.

LEBOT, V. **Tropical root and tuber crops: Cassava, sweet potato, yams and aroids**. Oxford: CAB International, 2009. p. 413.

LÉOTARD, G.; DUPUTIÉ, A.; KJELLBERG, F.; DOUZERY, E.J.P.; DEBAIN, C.; GRANVILLE, J.-J.D; McKEY, D. Phylogeography and the origin of cassava: New insights from the northern rim of the Amazonian basin. **Molecular Phylogenetics and Evolution**, Massachusetts, v. 53, n.1, p. 329-334, 2009.

LI, H.; HANDSAKER, B.; WYSOKER, A. et al. The sequence alignment/map (SAM) format and SAMtools. **Bioinformatics**, Oxford, v.25, n.16, p. 2078–2079, 2009.

LIN, Z.; LI, X.; SHANNON, L.M.; YEH, C.T.; WANG, M.L. BAI, G.; PENG, Z.; LI, J.; TRICK, H.N.; CLEMENTE, T.E.; DOEBLEY, J.; SCHNABLE, P.S.; TUINSTRA, M.R.; TESSO, T.T.; WHITE, F.; YU, J. Parallel domestication of the *Shattering1* genes in cereals. **Nature Genetics**, New York, v. 44, n. 6, p. 720-724, 2012.

LOUETTE, D.; CHARRIER, A.; BERTHAUD, J. *In situ* conservation of maize in Mexico: genetic diversity and maize seed management in a traditional community. **Economic Botany**, New York, v. 51, n. 1, p. 20-38, 1997.

LUIKART, G. H.; ENGLAND, P.; TALL LUIKART, G. H.; ENGLAND, P.; TALLMON, D. A.; JORDAN, S.; TABERLET, P. The power and promise of population genomics: from genotyping to genome-typing. **Nature Reviews Genetics**, London, v. 4, p. 981-994, 2003.

MARTINS, P.S. Dinâmica evolutiva em roças de caboclos amazônicos. In: VIEIRA, I.C.G.; SILVA, J.M.C.; OREN, D.C.; D'INCAO, M.A. (Ed). **Diversidade biológica e cultural da Amazônia**. Belém: Museu Paraense Emílio Goeldi, 2001. p. 369-384.

MBA, R.E.C.; STEPHENSON, P.; EDWARDS, K.; MLEZER, S.; MKUMBIRA, J.; GULLBERG, U.; APEL, K.; GALE, M.; TOHME, J.; FREGENE, M. Simple sequence repeat (SSR) markers survey of the Cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. **Theoretical and Applied Genetics**, Berlin, v. 102, p. 21-31, 2001.

McKEY, D.; BECKERMAN, S. Chemical ecology, plant evolution and traditional Manioc cultivation systems. In: HLADIK, C.M.; LINARES, O.F.; PAGEZY, H.; SEMPLÉ, A.; HADLEY, M. (Ed). **Tropical Forests, People and Food Biocultural Interactions and Applications to Development**. Paris: Parthenon Carnforth and UNESCO, 1993. v. 13, p. 83-

112.

McKEY, D.; CAVAGNARO, T.R.; CLIFF, J.; GLEADOW, R. Chemical ecology in coupled human and natural systems: people, manioc, multitrophic interactions and global change. **Chemoecology**, New York, v. 20, n. 2, p.109-133, 2010b.

McKEY, D.; ELIAS, M.; PUJOL, B.; DUPUTIÉ, A. Ecological approaches to crop domestication. In: GEPTS, P.; FAMULA, T. R.; BETTINGER, R.; BRUSH, S.B.; DAMANIA, A.B.; McGUIRE, P.E.; QUALSET C.O. (Ed). **Biodiversity in Agriculture: Domestication, Evolution and Sustainability**. Cambridge: Cambridge University Press, 2012. p. 377-406.

McKEY, D.; ELIAS, M.; PUJOL, B.; DUPUTIÉ, A. The evolutionary ecology of clonally propagated domesticated plants. **New Phytologist**, Lancaster, v. 186, n. 2, p.318-332, doi:10.1111/j.1469-8137.2010.03210.x. 2010a.

MEYER, R.S.; PURUGGANAN, M.D. Evolution of crop species: genetics of the domestication and diversification. **Nature Reviews Genetics**, London, v.14, p. 840-852, 2013.

MICHALAKIS, Y.; EXCOFFIER, L.G.L. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. **Genetics**, Austin, v. 142, p.1061-1064, 1996.

MIGNONE, F.; GISSI, C.; LIUNI, S.; PESOLE, G. Untranslated regions of mRNAs. **Genome Biology**, London, v. 3, n. 3, doi:10.1186/gb-2002-3-3-reviews0004, 2002.

MILACH, S.C.K. **Marcadores moleculares em plantas**. Porto Alegre: UFRGS, 1998. p. 140.

MILLER, M.R.; DUNHAM, J.P.; AMORES, A.; CRESKO, W.A.; JOHNSON, E.A. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. **Genome Research**, New York, v. 17, p. 240-248, 2007.

MILLER, A.; SCHAAL, B. Domestication of a Mesoamerican cultivated fruit tree, *Spondias purpurea*. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 102, n. 6, p. 12801-12806, 2005.

MONDINI, L.; NOORANI, A.; PAGNOTTA, M.A. Assessing plant genetic diversity by molecular tools. **Diversity**, Basel, v.1, p. 19-35, 2009.

MORIN, P.A.; LUIKART, G.; WAYNE, R.K.; SNP_WORKSHOP_GROUP. SNPs in ecology, evolution and conservation. **Trends in Ecology and Evolution**, Cambridge, v. 19, p. 208-216, 2004.

MÜHLEN, G.S.; ALVES-PEREIRA, A.; CLEMENT, C.R.; VALLE, T.L. Genetic diversity and differentiation of Brazilian bitter and sweet manioc varieties (*Manihot esculenta* Crantz, Euphorbiaceae) based on SSR molecular markers. **Tipiti**, v. 11, n. 2, p. 66-73, 2013.

MÜHLEN, G.S.; MARTINS, P.S.; ANDO, A. Variabilidade genética de etnovarietades de

mandioca, avaliada por marcadores de DNA. **Scientia Agricola**, Piracicaba, v. 57, n. 2, p. 319-328, 2000.

NASS, L.L. Utilização de recursos genéticos vegetais no melhoramento. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-INGLIS, M.C. (Ed.). **Recursos genéticos & melhoramento - plantas**. Rondonópolis: Fundação de Apoio à Pesquisa Agropecuária de Mato Grosso, 2001. p. 29-56.

NEI, M. **Molecular Evolutionary Genetics**. New York, Columbia University Press, 1987. p. 512.

NELSON, D.L.; COX, M.M. **Princípios de bioquímica de Lehninger**. Porto Alegre, ArtMed, 2011. p.1274.

NIELSEN, R.; WILLIAMSON, S.; KIM, Y.; HUBISZ, M.J.; CLARK, A.G.; BUSTAMANTE, C. Genomic scans for selective sweeps using SNP data. **Genome Research**, New York, v.15, p.1566-1575, 2005.

OLIVEIRA, E.J.; FERREIRA, C.F.; SANTOS, V.S.; JESUS, O.N.; OLIVEIRA, G.A.F.; SILVA, M.S. Potential of SNP markers for the characterization of Brazilian cassava germplasm. **Theoretical and Applied Genetics**, Berlin, v. 127, p. 1423-1440, 2014.

OLIVEIRA, E.J.; RESENDE, M.D.V.; SANTOS, V.S.; FERREIRA, C.F.; OLIVEIRA, G.A.F.; SILVA, M.S.; OLIVEIRA, L.A.; AGUILAR-VILDOSO, C.I. Genome-wide selection in cassava. **Euphytica**, Wageningen, v. 187, n. 2, p. 263-276, 2012.

OLIVEIRA, J.C. Social networks and cultivated plants. **Tipití**, v. 6, n. 1-2, p.101-110, 2008.

OLIVER, J.R. The archaeology of agriculture in ancient Amazonia. In: SILVERMAN, H.; ISBELL, W. (Ed). **Handbook of South American Archaeology**. New York: Springer, 2008. p.185-216.

OLSEN, K.M. SNPs, SSRs and inferences on Cassava's origin. **Plant Molecular Biology**, New York, v. 56, p. 517-526, 2004.

OLSEN, K.M.; SCHAAL, B.A. Evidence on the origin of Cassava: Phylogeography of *Manihot esculenta*. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 96, p. 5586-5591, 1999.

_____. Microsatellite variation in Cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: Further evidence for a Southern Amazonian origin of domestication. **American Journal of Botany**, Baltimore, v. 88, n. 1, p. 131-142, 2001.

OUBORG, N.J.; PERTOLDI, C.; LOESCHCKE, V.; BIJLSMA, R.; HEDRICK, P.W. Conservation genetics in transition to conservation genomics. **Trends in Genetics**, London, v. 26, n. 4, p. 177-187, 2010.

PEAKALL, R.; SMOUSE, P.E. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. **Bioinformatics**, Oxford, v.28, p. 2537-2539, 2012.

PEGADARAJU, V.; NIPPER, R.; HULKE, B.; QI, L.; SCHULTZ, Q. *De novo* sequencing of sunflower genome for SNP discovery using RAD (Restriction site Associated DNA) approach. **BMC Genomics**, Austin, v. 14:556, doi:10.1186/1471-2164-14-556, 2013.

PEÑA-VENEGAS, C.P.; STOMPH, T.J.; VERSCHOOR, G.; LOPEZ-LAVALLE, L.A.B.; STRUIK, P.C. Differences in manioc diversity among five ethnic groups of Colombia Amazon. **Diversity**, Basel, v. 6, p. 792-826, 2014.

PETERSON, B.K.; WEBER, J.N.; KAY, E.H.; FISHER, H.S.; HOEKSTRA, H.E. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. **PLoS ONE**, San Francisco, v. 7, n. 5, e37135, doi:10.1371/journal.pone.0037135, 2012.

PETIT, R.J.; DUMINIL, J.; FINESCHI, S.; HAMPE, A.; SALVINI, D.; VENDRAMIN, G.G. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. **Molecular Ecology**, Oxford, v. 14, p. 689-701, 2005.

PERONI, N. **Taxonomia folk e diversidade intraespecífica de mandioca (*Manihot esculenta* Crantz) em roças de agricultura tradicional em Áreas de Mata Atlântica do Sul do Estado de São Paulo**. 1998. Dissertação de Mestrado (Genética e Melhoramento de Plantas), Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1998.

_____. **Ecologia e genética da mandioca na agricultura itinerante do litoral sul paulista: Uma análise espacial e temporal**. 2004. 246p. Tese de Doutorado (Biologia Vegetal), Universidade Estadual de Campinas, São Paulo, 2004.

PERONI, N.; HANAZAKI, N. Current and lost diversity of cultivated varieties, especially Cassava, under swidden cultivation systems in the Brazilian Atlantic Forest. **Agriculture, Ecosystems and Environment**, Amsterdam, v. 92, n. 2-3, p. 171-183, 2002.

PERONI, N.; KAGEYAMA, P.; BEGOSSI, A. Molecular differentiation, diversity, and folk classification of "sweet" and "bitter" Cassava (*Manihot esculenta*) in Caçara and Caboclo management systems (Brazil). **Genetic Resources Crop Evolution**, Dordrecht, v. 54, p. 1333-1349, 2007.

PERRUT-LIMA, P.; MÜHLEN, G.S.; CARVALHO, C.R.L. Cyanogenic glycoside content of *Manihot esculenta* subsp. *flabellifolia* in south-central Rondônia, Brazil, in the center of domestication of *M. esculenta* subsp. *esculenta*. **Genetic resources and Crop Evolution**, Dordrecht, DOI 10.1007/s10722-014-0105-6, 2014.

PERRY, L. Starch granule size and the domestication of manioc (*Manihot esculenta*) and sweet potato (*Ipomoea batatas*). **Economic Botany**, New York, v. 56, n. 4, p. 335-349, 2002.

PICKERSGILL, B. Domestication of plants in the Americas: insights from Mendelian and molecular genetics. **Annals of Botany**, Oxford, v. 100, p. 925-940, 2007.

PIPERNO, D.R.; HOLST, I. The presence of starch grains on prehistoric stone tools from the humid Neotropics: indications of early tuber use and agriculture in Panama. **Journal of**

Archaeological Science, v. 25, n. 8, p. 765-776, 1998.

POOTAKHAM, W.; SHEARMAN, J.R.; RUANG-AREERATE, P.; SONTIHIROD, C.; SANGSRAKRU, D.; JOMCHAI, N.; YOOCHA, T.; TRIWITAYAKORN, K.; TRAGOONRUNG, S.; TANGPHATSORNRUANG, S. Large-scale SNP discovery through RNA sequencing and SNP genotyping by targeted enrichment sequencing in cassava (*Manihot esculenta* Crantz). **PLoS ONE**, San Francisco, v. 9, n.12, e116028, doi:10.1371/journal.pone.0116028, 2014.

POWELL, W.; MACHRAY, G.C.; PROVAN, J. Polymorphism revealed by simple sequence repeats. **Trends in Plant Science**, London, v. 1, n. 7, p. 215-222, 1996.

POWELL, W.; MORGANTE, M.; McDEVITT, R.; VENDRAMIN, G.G.; RAFALSKI, J.A. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 92: p. 7759–7763, 1995a.

POWELL, W.; MORGANTE, M.; ANDRE, C.; MCNICOL, J.W.; MACHRAY, G.C.; DOYLE, J.J.; TINGEY, S.V.; RAFALSKI, J.A. Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. **Current Biology**, Chicago, v. 5, p. 1023–1029, 1995b.

PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, Austin, v. 155, p. 945-959, 2000.

PROCHNIK, S.; MARRI, P.R.; DESANY, B.; RABINOWICZ, P.D.; KODIRA, C.; MOHIUDDIN, M.; RODRIGUEZ, F.; FAUQUET, C.; TOHME, J.; HARKINS, T.; ROKHSAR, D.S.; ROUNSLEY, S. The cassava genome: current progress, future directions. **Tropical Plant Biology**, New York, v. 5, p.88-94, 2012.

PROVAN, J.; POWELL, W.; HOLLINGSWORTH, P.M. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. **Trends in Ecology & Evolution**, Cambridge, v. 16, n. 3, p. 142-147, 2001.

PUJOL, B.; McKEY, D. Size asymmetry in intraspecific competition and the density dependence of inbreeding depression in a natural plant population: a case study in cassava (*Manihot esculenta* Crantz, Euphorbiaceae). **Journal of Evolutionary Biology**, Oxford, v.19, p. 85-96, 2006.

PUJOL, B.; DAVID, P.; McKEY, D. Microevolution in agricultural environments: How a traditional Amerindian farming practice favours heterozygosity in Cassava (*Manihot esculenta* Crantz, Euphorbiaceae). **Ecology Letters**, Oxford, v. 8, p. 138-147, 2005.

PUJOL, B.; GIGOT, G.; LAURENT, G.; PINHEIRO-KLUPPEL, M.; ELIAS, M.; HOSSAERT-McKEY, M.; McKEY, D. Germination ecology of Cassava (*Manihot esculenta* Crantz, Euphorbiaceae) in traditional agroecosystems: Seed and seedling biology of a vegetatively propagated domesticated plant. **Economic Botany**, New York, v. 56, n. 4, p. 366-379, 2002.

PUJOL, B.; MÜHLEN, G.; GARWOOD, N.; HOROSZOWSKI, Y.; DOUZERY, E.J.P.;

McKEY, D. Evolution under domestication: contrasting functional morphology of seedlings in domesticated cassava and its closest wild relatives. **New Phytologist**, Lancaster, v. 166, p. 305-318, 2005.

PUJOL, B.; RENOUX, F.; ELIAS, M.; RIVAL, L.; McKEY, D. The unappreciated ecology of landrace populations: Conservation consequences of soil seedbanks in Cassava. **Biological Conservation**, Boston, v. 136, p. 541-551, 2007.

PURITZ, J.B.; MATZ, M.V.; TOONEN, R.J.; WEBER, J.N.; BOLNICK, D.I.; BIRD, C.E. Demystifying the RAD fad. **Molecular Ecology**, Oxford, v. 23, p. 5937-5942, 2014.

PURUGGANAN, M.D.; FULLER, D.Q. The nature of selection during plant domestication. **Nature**, London, v.47, p. 843-848, 2009.

R Development Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2003. Disponível em: < <http://www.R-project.org>>. Acesso em: 15 de janeiro de 2013.

RIVAL, L.; McKEY, D. Domestication and diversity in Manioc (*Manihot esculenta* Crantz ssp. *esculenta*, Euphorbiaceae). **Current Anthropology**, Chicago, v. 49, n. 6, p. 1119-1128, 2008.

ROCHA, O.J.; ZALDIVAR, M.E.; CASTRO, L.; CASTRO, E.; BARRANTES, R. Microsatellite variation of cassava (*Manihot esculenta* Crantz) in home gardens of Chibchan Amerindians from Costa Rica. **Conservation Genetics**, Arlington, v. 9, p. 107-118, 2008.

ROGERS, D.J.; APAN, S.G. 1973. **Manihot, Manihotoides (Euphorbiaceae)**. In: Flora Neotropica (Monograph No. 13). Hafner Press, New York.

ROSENBERG, N.A. Distruct: a program for the graphical display of population structure. **Molecular Ecology Notes**, Oxford, v. 4, p. 137-138, 2004.

ROULLIER, C.; BENOIT, L.; McKEY, D.B.; LEBOT, V. Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v.110, n. 6, p. 2205-2210, 2013.

ROULLIER, C.; ROSSEL, D.; TAY, D.; McKEY, D.B.; LEBOT, V. Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. **Molecular Ecology**, Oxford, v.20, p. 3963-3977, 2011.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, Chicago, v. 4, n. 4, p. 406-425, 1987.

SALICK, J.; CELLINESE, N.; KNAPP, N. Indigenous diversity of cassava: Generation, maintenance, use and loss among the Amuesha, Peruvian Upper Amazon. **Economic Botany**, New York, v. 51, n. 1, p. 6-17, 1997.

SAMBATTI J.B.M.; MARTINS, P.S.; ANDO, A. Folk taxonomy and evolutionary dynamics of cassava: A case study in Ubatuba, Brazil. **Economic Botany**, New York, v. 55, n. 1, p. 93-

105, 2001.

SARDOS, J.; McKEY, D.; DUVAL, M.F.; MALAPA, R.; NOYER, J.L.; LEBOT, V. Evolution of cassava (*Manihot esculenta* Crantz) after recent introduction into a South Pacific Island system: the contribution of sex to the diversification of a clonally propagated crop. **Genome**, Ottawa, v. 51, p. 912-921, 2008.

SCHELDEMAN, X.; vanZONNEVELD, M. **Training manual on spatial analysis of plant diversity and distribution**. Rome, Bioversity International, 2010. p. 179.

SEEB, J.E.; CARVALHO, G.; HAUSER, L.; NAISH, K.; ROBERTS, S.; SEEB, W. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. **Molecular Ecology Resources**, Oxford, v. 11, suppl. 1, p.1-8, 2011.

SCHAAL, B.A.; HAYWORTH, D.A.; OLSEN, K.M.; RAUSCHER, J.T.; SMITH, W.A. Phylogeographic studies in plants: problems and prospects. **Molecular Ecology**, Oxford, v. 7, p. 465-474, 1998.

SCHAAL, B.A.; OLSEN, K.M.; CARVALHO, L.J.C.B. Evolution, domestication, and agrobiodiversity in the tropical crop cassava. In: MOTLEY, T.J.; ZEREGA, N; HUGH, H. (Ed). **Darwin's harvest – New approaches to the origins, evolution, and conservation of crops**. New York: Columbia University Press, 2006. p.269-284.

SCHLÖTTERER, C. The evolution of molecular markers – just a matter of fashion? **Nature Reviews Genetics**, London, v. 5, p. 63-69, 2004.

SIQUEIRA, M.V.B.M.; QUEIROZ-SILVA, J.R.; BRESSAN, E.; BORGES, A.; PEREIRA, K.J.C; PINTO, J.G.; VEASEY, E.A. Genetic characterization of Cassava (*Manihot esculenta*) landraces in Brazil assed with simple sequence repeats. **Genetics and Molecular Biology**, Ribeirão Preto, v. 32, n. 1, p. 104-110, 2009.

SKINNER, M.E.; UZILOV, A.V.; STEIN, L.D.; MUNGALL, C.J.; HOLMES, I.H. JBrowse: a next-generation genome browser. **Genome Research**, New York, v. 19, p. 1630-1638, 2009.

SMITH, R.N.; ALEKSIC, J.; BUTANO, D.; CARR, A.; CONTRINO, S.; HU, F.; LYNE, M.; LYNE, R.; KALDERIMIS, A.; RUTHERFORD, K. et al. InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. **Bioinformatics**, Oxford, v. 28, p. 3163–3165, 2012.

STÖLTING, K.N.; NIPPER, R.; LINDTKE, D.; CASEYS, C.; WAEBER, S.; CASTIGLIONE, S.; LEXER, C. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. **Molecular Ecology**, Oxford, v. 22, p. 842-855, 2013.

STUDER, A.; ZHAO, Q.; ROSS-IBARRA, J.; DOEBLEY, J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. **Nature Genetics**, New York, v. 43, p. 1160-1163, 2011.

TANG, H.; SEZEN, Y.; PATERSON, A.H. Domestication and plant genomes. **Current Opinion in Plant Biology**, London, v. 13, p. 160-166, 2010.

TAUTZ, D.; ELLEGREN, H.; WEIGEL, D. Next generation molecular ecology. **Molecular Ecology**, Oxford, v. 19, suppl. 1, p. 1-3, 2010.

TOONEN, R.J.; PURITZ, J.B.; FORSMAN, Z.H.; WHITNEY, J.L.; FERNANDEZ-SILVA, I.; ANDREWS, K.R.; BIRD, C.E. ezRAD: a simplified method for genomic genotyping in non-model organisms. **PeerJ**, v. 1, e203, doi:10.7717/peerj.203, 2013.

UGENT, D. Archaeological manioc (*Manihot*) from coastal Peru. **Economic Botany**, New York, v. 40, n. 1, p. 78-102.

VanZONNEVELD, M.; SCHELDAMAN, X.; ESCRIBANO, P.; VIRUEL, M.A.; VanDAMME, P., et al. Mapping Genetic Diversity of Cherimoya (*Annona cherimola* Mill.): Application of Spatial Analysis for Conservation and Use of Plant Genetic Resources. **PLoS ONE**, San Francisco, v. 7, n. 1, e29845, doi:10.1371/journal.pone.0029845, 2012.

VENCOVSKY, R.; NASS, L.L.; CORDEIRO, C.M.T.; FERREIRA, M.A.J.F. Amostragem em recursos genéticos vegetais. In: NASS, L.L. (Ed.). **Recursos genéticos vegetais**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2007. p. 231-280.

WANG, S.; MEYER, E.; McKAY, J.K.; MATZ, M.V. 2b-RAD: a simple and flexible method for genome-wide genotyping. **Nature Methods**, New York, v. 9, p. 808-810, 2012.

WEISING, K.; GARDNER, R.C. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. **Genome**, Ottawa, v. 42, p. 9-19, 1999.

WHITLOCK, M.C. G'_{ST} and D do not replace F_{ST} . **Molecular Ecology**, Oxford v. 20, n. 6, p. 1083-1091, 2011.

WHITLOCK, M.C.; McCAULEY, D.E. Indirect measure of gene flow and migration: F_{ST} is not equal to $1/(4Nm+1)$. **Heredity**, London, v. 82, p. 117-125, 1999.

WILLING, E.-M.; DREYER, C.; VAN OOSTERHOUT, C. Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many snps markers. **PLoS ONE**, San Francisco, v. 7, n. 8, e42649, DOI:10.1371/journal.pone.0042649, 2012.

WOODS, W.I.; DENEVAN, W.M. Amazonian dark earths: The first century of reports. In: WOODS, W.I.; TEIXEIRA, W.G.; LEHMANN, J.; STEINER, C.; WINKLERPRINS, A.M.G.A.; REBELLATO, L. (Ed). **Amazonian Dark Earths: Wim Sombroek's Vision**. New York: Springer, 2009. p. 1-13.

ZALAPA, J.; CUEVAS, H.; ZHU, H.; STEFFAN, S.; SENALIK, D.; ZELDIN, E.; MCCOWN, B.; HARBUT, R.; SIMON, P. Using next-generation sequencing approaches to isolate simple sequence repeats (SSR) loci in plant sciences. **American Journal of Botany**, Baltimore, v. 99, n. 2, p. 193-208, 2012.