



**MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA
DOUTORADO EM AGRONOMIA**

FRANCISCA DAS CHAGAS BEZERRA DE ARAÚJO

**ESTUDO DE ACESSOS DE MANDIOCA DO BANCO ATIVO DE GERMOPLASMA
DA AMAZÔNIA ORIENTAL POR MEIO DE DESCRITORES QUANTITATIVOS E
MARCADORES MOLECULARES SNP**

**BELÉM
2017**

FRANCISCA DAS CHAGAS BEZERRA DE ARAÚJO

**ESTUDO DE ACESSOS DE MANDIOCA DO BANCO ATIVO DE GERMOPLASMA
DA AMAZÔNIA ORIENTAL POR MEIO DE DESCRITORES QUANTITATIVOS E
MARCADORES MOLECULARES SNP**

Tese apresentada ao Programa de Pós-Graduação em
Agronomia da Universidade Federal de Rural da
Amazônia, como requisito para obtenção do título de
Doutorado em Agronomia.

Orientadora: Dra. Elisa Ferreira Moura Cunha

Co-orientador: Dr. Gabriel Rodrigues Alves Margarido

**BELÉM
2017**

Dados Internacionais de Catalogação de Publicação-na-Fonte (CIP)
Ficha Catalográfica

A658

Araújo, Francisca das Chagas Bezerra de

Estudo de acessos de mandioca do banco ativo de germoplasma da Amazônia oriental por meio de descritores quantitativos e marcadores moleculares SNP. / Francisca das Chagas Bezerra de Araújo.- Capitão Poço: [S.n], 2017.
91f.

Orientadora: Dr^a. Elisa Ferreira Moura Cunha
Programa de Pós-Graduação em Agronomia. (Doutorado em Agronomia). – Universidade Federal Rural da Amazônia.

1. Amido. 2. Caracterização química. 3. Carotenóides. 4. Genotipagem por sequenciamento. 5. Haplótipos. 5. Modelagem de proteína. 6. Polimorfismo.

CDD 633.682

Bibliotecária: Vanessa Baía CRB/2-1322

FRANCISCA DAS CHAGAS BEZERRA DE ARAÚJO

**ESTUDO DE ACESSOS DE MANDIOCA DO BANCO ATIVO DE GERMOPLASMA
DA AMAZÔNIA ORIENTAL POR MEIO DE DESCRITORES QUANTITATIVOS E
MARCADORES MOLECULARES SNP**

Tese apresentada ao Programa de Pós-Graduação em Agronomia, da Universidade Federal Rural da Amazônia, como requisito para obtenção do título de Doutora em Agronomia.

Orientadora: Dra. Elisa Ferreira Moura Cunha

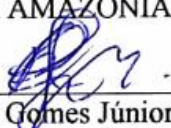
Co-orientador: Professor Dr. Gabriel Rodrigues Alves Margarido

Aprovada em 30 de junho de 2017


BANCA EXAMINADORA



Dra. Elisa Ferreira Moura Cunha – Presidente
EMBRAPA AMAZÔNIA ORIENTAL



Dr. Rui Alberto Gomes Júnior - 2º Examinador
EMBRAPA AMAZÔNIA ORIENTAL



Dra. Simone de Miranda Rodrigues - 3º Examinador
EMBRAPA AMAZÔNIA ORIENTAL



Dr. Didier Clement – 4º Examinador

CENTRE DE COOPÉRATION INTERNATIONALE EM RECHERCHE AGRONOMIQUE
POUR LE DÉVELOPPERMENT- CIRAD



Dr. Marcelo Murad Magalhães – Suplente
EMBRAPA AMAZÔNIA ORIENTAL

BELÉM

2017

AGRADECIMENTOS

A Deus, por estar em meu coração, onde eu encontro força para ir em busca dos meus sonhos e encarar os desafios da vida.

A Universidade Federal Rural da Amazônia e ao Programa de Pós-Graduação em Agronomia pela oportunidade de realizar este curso.

A Embrapa Amazônia Oriental pela disponibilização dos laboratórios onde eu realizei as análises.

A minha Orientadora doutora Elisa Ferreira Moura pela oportunidade da orientação, pelos conhecimentos compartilhados, paciência, dedicação, prontidão e serenidade durante a realização do curso de doutorado.

Ao meu co-orientador professor Gabriel Rodrigues Alves Margarido por ter me recebido em sua equipe de trabalho e pela minha permanência no Laboratório de Bioinformática Aplicada a Bioenergia- ESALQ. Gabriel você é um exemplo de profissional, obrigada pelos ensinamentos, paciência e por ter me apresentado o Linux e todos os softwares que jamais imaginei que aprenderia usar.

Aos técnicos e colegas do Laboratório de Genética da Embrapa Amazônia Oriental, por me auxiliar nas análises de extração do DNA. Agradeço especialmente as amigas Ilenilce, Lígia, Andréia e Leonária e aos amigos Pedro, Isaías e Caio, vocês fizeram meus dias mais felizes.

A Dra. Kenny Bonfim e Mônica Fercury e ao professor Davidson pelos ensinamentos nas análises e pela amizade.

As Dra. Ilmarina Campos e Simone Miranda, pelos ensinamentos e pelo apoio durante minha permanência na Embrapa.

Ao Dr. Gilvan pela paciência e ensinamentos, Gilvan muito obrigada por sempre estar a disposição para colaborar em meu doutorado, mesmo não nos conhecendo pessoalmente você sempre foi muito gentil.

Aos colegas do Laboratório de Bioinformática Aplicada a Bioenergia pelos ensinamentos de bioinformática e genética. Agradeço especialmente a Lorena Batista, Leonardo, Amanda, Fernando e Guilherme pela atenção e ensinamentos, vocês foram anjos em minha vida.

Ao Dr. Eder Jorge de Oliveira que articulou o envio das amostras para o sequenciamento.

Ao meu amigo Rodrigo Souza, pela colaboração e ensinamentos em estatística.

Ao Dr. Eraldo Tavares pela contribuição nas análises de modelagem.

A banca avaliadora deste trabalho: Dr. Rui Alberto Gomes Júnior, Dra. Simone de Miranda Rodrigues, Dr. Didier Clement e Dr. Marcelo Murad Magalhães, pelas contribuições.

Ao meu esposo Fábio pela compreensão, companheirismo, amor e apoio incondicional durante todos os momentos da minha vida.

Aos meus pais Anastácio e Terezinha pelas orações e incentivo nos meus estudos. Pai e mãe vocês são meu maior exemplo de perseverança e fé. A minha irmã Flávia e minha sogra Sueli pelo incentivo, apoio e orações.

As amigas Ana Carolina Sonsin, Edna, Aline Medeiro, Suely Cristina, Lucila, Amanda, Carolina Melo e Jackeline e aos amigos Jonny Lúcio, Alberto, Robson e Fadul pelo companheirismo nos momentos difíceis da vida.

A todos que de alguma forma contribuíram para a realização deste trabalho, sintam-se eternamente agradecidos.

MUITO OBRIGADA!

DEDICATÓRIA

“... Ao meu amado esposo Fábio, aos meus pais Anastácio Ximendes e Terezinha Bezerra e a minha irmã Flávia que me apoiaram incansavelmente nesta jornada...”

LISTA DE ILUSTRAÇÕES

- Figura 1.1.** Passos para a construção de bibliotecas de GBS 20
- Figura 2.1.** Representação esquemática do gene fitoeno sintase 2 (*psy2*) de mandioca (*Manihot esculenta*), mostrando as regiões de ligação dos primers forward e reverse. O éxon está destacado em cinza, as regiões de ligação dos primers estão destacadas em vermelho e as letras pretas sem preenchimento pertencem a uma região de íntron. 36
- Figura 2.2.** Rede de haplótipos de acessos de mandioca (*Manihot esculenta* Crantz) com tonalidade branca, creme e amarela. **a** Rede de haplótipo construída com base nas sequências de 173 pb do éxon 3 do gene da fitoeno sintase 2 (*psy2*) de 78 acessos de mandioca obtida pelo programa Haploview. A cor cinza dos círculos representa acessos com raiz de tonalidade branca, cor amarela dos círculos representa acessos com raiz de tonalidade amarela ou creme. O tamanho do círculo é equivalente ao tamanho da amostra. **b** Representação das raízes de mandioca branca, creme e amarela. 40
- Figura 2.3.** Rede de haplótipos construída com base nas sequências de aminoácidos do éxon 3 do gene que codificadora a proteína fitoeno sintase 2 (*psy2*) com base em 78 acessos de mandioca da região Amazônica. A rede de haplótipo foi obtida pelo programa Haploview. O círculo cinza apresenta haplótipos com raiz de tonalidade branca, os círculos amarelos apresentam haplótipos com raiz de tonalidade amarela e creme. O tamanho do círculo é equivalente ao tamanho da amostra. 41
- Figura 2.4.** Variantes da proteína e do gene da fitoeno sintase 2 em acessos de mandioca (*Manihot esculenta*) com raízes de diferentes tonalidades. **a** Alinhamento das sequências de aminoácidos de acessos com tonalidade de raiz branca (W), creme (C) e amarela (Y), retângulo verde acessos com homozigose para o SNP1, retângulo de cor vermelha representa acesso com homozigose para o SNP3. **b** Picos do cromatograma de SNPs em homozigose e heterozigose da região funcional do gene fitoeno sintase 2 (*psy2*). 43
- Figura 2.5.** Distribuição geográfica dos oito haplótipos do gene fitoeno sintase 2 (*psy2*) de mandioca (*Manihot esculenta* Crantz) identificados após o sequenciamento de fragmento de 300 pb dos 78 acessos originários da Amazônia brasileira. 44
- Figura 2.6.** Modelos preditivos de fitoeno sintase. Para maior clareza, apenas os resíduos 133-416 são mostrados. Os resíduos mais conservados entre a família de esqualeno e fitoeno sintase, bem como a região importante 146-149, e os mutantes 174 como 191 são coloridos pelo nome do resíduo. Os arredores são coloridos por cadeias. **a** Uma representação em 3D do acy42665.1. Os resíduos conservados formam um núcleo com uma cavidade, onde o sítio ativo presumido está localizado. Os resíduos 174 e 191 foram destacados com o raio Van der Waals e parecem estar separados da superfície da cavidade. **b** Um alinhamento 3D dos quatro mutantes, usando acy42665.1 como referência. Os dois motivos ricos em Asp₁₇₆DELVD₁₈₀ e₃₀₂DVGED₃₀₆ se enfrentam na região externa do canal. Os resíduos 146-150 correspondem à "flap" flexível de SQS- Esqualeno sintase (resíduos 50-54), que se pensa proteger da água os intermediários de reação (Umeno et al 2002, Pandit et al., 2000). 45
- Figura 2.7.** O RMSD por resíduo entre os mutantes dos acessos. Acy42665.1 é a referência. Os resíduos 174 e 191 são os mutantes, e na sequência de referência são R174 e A191. Os valores mais altos tendem a estar dentro das regiões com mais características de alças. Os resíduos Asp D176, D180, D302 e D306 são altamente conservados. A mutação do resíduo 302 inativa a enzima. 46

Figura 3.1. Dispersão de 398 acessos com base na análise de componentes principais. CP 1 componente principal 1, CP 2 componente principal 2.....	68
Figura 3.2. Distribuição da frequência do alelo alternativo (MAF- <i>minor allele frequency</i>), com base nos dados de genotipagem não filtrados.....	69
Figura 3.3. Distribuição da frequência do conteúdo de amido na raiz de mandioca.....	70
Figura 3.4. Gráfico quantil-quantil dos valores de $-\log_{10}(p)$ esperado contra o observado para a característica percentual de amido.....	71
Figura 3.5. Manhattan plot da análise de associação genômica entre $-\log_{10}$ (p-valor) e o loci do SNP em ordem sequencial por cromossomo, para característica conteúdo de amido.....	72
Figure 4.1. Graphical dispersion of 46 genotypes of cassava (<i>Manihot esculenta</i> Crantz) sampled in the Amazon and analyzed with 13 chemical traits of the root based on the first two principal components. Scaling method used: correlation matrix.	86
Figure 4.2. Dendrogram generated by UPGMA method based on standardized Euclidean mean distances as measure of dissimilarity, and obtained for 46 genotypes of cassava (<i>Manihot esculenta</i> Crantz) sampled in the Amazon and analyzed with 13 chemical traits of the root. Coefficient of cophenetic correlation ($r = 0.71$). Groups were determined based on ‘NbClust’ package of software R 3.0.3 (R Core Team) using Pseudot2. (S) indicates ‘sweet’ genotypes of cassava.	87

LISTA DE TABELAS

Tabela 2.1. Informações dos acessos de mandioca (<i>Manihot esculenta</i> , Crantz) pertencentes ao Banco Ativo de Germoplasma (BAG) da Amazônia Oriental, localizado em Belém, Pará, Brasil.....	32
Tabela 2.2. Identificação dos SNPs na região do éxon 3 do gene da fitoeno sintase 2 (<i>psy2</i>) e em 78 acessos de mandioca da Amazônia e seus respectivos haplótipos.	38
Tabela 3.1 Classificação ascendente dos 10 menores P-valores para o fenótipo conteúdo de amido da raiz de mandioca, gerados a partir do MLM e utilizados na análise de anotações dos SNP.....	67
Table 4.1. Descriptive analysis of 13 chemical traits of the root of 46 genotypes of cassava (<i>Manihot esculenta</i> Crantz) sampled in the Brazilian Amazon and maintained in a germplasm bank from Eastern Amazon, Brazil. Minimum (Min.), maximum (Max.) and mean values, amplitude, standard deviation, coefficient of variation (CV %), and normality test.....	81
Table 4.2. Chemical traits of the root of genotypes of ‘bitter’ and ‘sweet’ cassava (<i>Manihot esculenta</i> Crantz) sampled in the Brazilian Amazon and that belong to the germplasm bank in Eastern Amazon, Brazil. Traits were measured on humid basis of roots.....	82
Table 4.3. Chemical traits of the root of genotypes of ‘bitter’ and ‘sweet’ cassava (<i>Manihot esculenta</i> Crantz) sampled in the Brazilian Amazon and that belong to the germplasm bank in Eastern Amazon, Brazil. Traits were measured on humid basis of roots.....	83
Table 4.4. Estimatives of correlation coefficient of Spearman among 13 chemical traits of the root measured in 46 genotypes of cassava (<i>Manihot esculenta</i> Crantz) sampled in the Brazilian Amazon and maintained in a germplasm bank in the Eastern Amazon, Brazil.....	84
Table 4.5. Estimatives of eigenvalues and eigenvectors associated with the principal components and 13 chemical traits of the root of cassava (<i>Manihot esculenta</i> Crantz) obtained for 46 genotypes sampled in the Brazilian Amazon and maintained in a germplasm bank in Eastern Amazon, Brazil.....	85

SUMÁRIO

1	CONTEXTUALIZAÇÃO	15
1.1	REFERÊNCIAS.....	21
2	IDENTIFICAÇÃO DE SNP NO GENE FITOENO SINTASE 2 (<i>psy2</i>) ASSOCIADOS A COLORAÇÃO AMARELA DA RAIZ EM ACESSOS DE MANDIOCA DA AMAZÔNIA BRASILEIRA	29
	RESUMO.....	29
	ABSTRACT	30
2.1	Introdução	30
2.2	Material e métodos.....	32
2.2.1	Material vegetal	32
2.2.2	Extração e quantificação do DNA total	34
2.2.3	Reação em cadeia da polimerase (PCR).....	35
2.2.4	Purificação e sequenciamento do DNA	36
2.2.5	Análise de bioinformática.....	36
2.2.6	Modelagem molecular	37
2.3	Resultados.....	38
2.4	Discussão	47
2.5	Conclusão.....	50
	Referências	51
3	ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA PARA CONTEÚDO DE AMIDO EM ACESSOS DE MANDIOCA DO BANCO DE GERMOPLASMA DA AMAZÔNIA ORIENTAL.....	55
	RESUMO.....	55
	ABSTRACT	55
3.1	Introdução	56
3.2	Materiais e métodos.....	57

3.2.1	Material vegetal	57
3.2.2	Fenotipagem.....	58
3.2.3	Genotipagem.....	58
3.2.4	Análises Estatísticas.....	59
3.3	Resultados e discussão.....	60
	REFERÊNCIAS	63
4	CHEMICAL ROOT TRAITS DIFFERENTIATE GENOTYPES OF ‘BITTER’ AND ‘SWEET’ CASSAVA (<i>Manihot esculenta</i>) SAMPLED IN THE AMAZON	73
	ABSTRACT	73
4.1	Introduction.....	73
4.2	Material and methods.....	74
4.3	Results and discussion	75
	REFERENCES	79
	CONCLUSÃO GERAL	88
	APPENDIX	89

RESUMO

A mandioca (*Manihot esculenta* Crantz) é classificada como a terceira maior fonte de carboidratos do mundo, consumida por mais de 800 milhões de pessoas e cultivada em países da África tropical, Ásia e América do sul. Nos países onde é consumida, o desenvolvimento de acessos ricos em nutrientes e alto conteúdo de amido é importante para prevenir doenças, melhorar a dieta e a aplicação industrial do amido. O estudo de acessos de mandioca mantidos em bancos de germoplasma é importante para apoiar programas de melhoramento genético, a fim de identificar acessos promissores que possam gerar cultivares biofortificadas e com elevado conteúdo de amido nas raízes. Para o estudo de acessos de mandioca mantidos no Banco Ativo de Germoplasma da Amazônia Oriental, foi realizado um estudo utilizando descritores quantitativos e marcador molecular SNP (*Single Nucleotide Polymorphism*). Ao avaliar 78 acessos de mandioca para verificar a presença de SNP em uma região de aproximadamente 300 pb no gene fitoeno sintase 2 (*psy2*), importante na síntese de carotenóides e associados a coloração da raiz, foram identificados três SNPs localizados na região codificante, e a existência de oito haplótipos. As variações c.1434G>C e c.1485C>A dos SNP1 e SNP3 foram encontrados em acessos de raiz de coloração amarela e creme respectivamente, e o polimorfismo de nucleotídeo resultou na substituição de aminoácidos. O efeito das mutações na estrutura e função da referida proteína também foi avaliado *in silico* pela dinâmica molecular. Os resultados foram compatíveis com o fenótipo detectado experimentalmente. Também foi utilizada a plataforma de genotipagem por sequenciamento para genotipar um painel de 398 acessos, onde foram identificados 80.327 SNPs distribuídos pelo genoma da mandioca. Dentro de uma região genômica do cromossomo 1, foi identificado um SNP associado ao conteúdo de amido na mandioca. Esse SNP foi posicionado na proximidade do locus Manes.01G048600 e a anotação funcional deste gene relaciona a sua atividade biológica com a enzima glucose-6-phosphate 1-epimerase, a qual é fortemente ligada ao metabolismo de carboidrato. Portanto, esses SNPs podem representar um importante marcador para seleção assistida em programas de melhoramento genético visando o aumento do conteúdo de amido e a biofortificação das raízes. Além disso, foi realizada a caracterização química de 46 acessos de mandioca "brava" e "doce" para verificar as relações genéticas entre acessos com base nas características químicas. Foi avaliada: umidade, cinzas, sólidos solúveis totais, acidez titulável total, pH, carotenóides totais, cianeto livre e total, proteína bruta, glicose, frutose, sacarose e amido. Foi detectada uma variação fenotípica considerável para os caracteres. Além do cianeto livre e total, características como açúcares e umidade

apresentaram diferentes médias para mandioca "brava" e "doce". O amido e as concentrações de cianeto livre e total foram os caracteres menos importantes para discriminar os acessos. As mandiocas "bravas" e "doces" foram agrupadas em diferentes clusters no dendrograma e na dispersão da análise de componentes principais. O banco de germoplasma da Amazônia Oriental apresentou variação genética para características químicas da raiz da mandioca.

Palavras-chaves: Amido. Caracterização química. Carotenóides. Genotipagem por sequenciamento. Haplótipos. Modelagem de proteína. Polimorfismo.

ABSTRACT

Cassava (*Manihot esculenta* Crantz) is classified as the third largest source of carbohydrates in the world, consumed by more than 800 million people and grown in countries of tropical Africa, Asia and South America. In countries where it is consumed, the development of access rich in nutrients and high starch content is important to prevent diseases, improve diet and industrial application of starch. The study of cassava accesses kept in germplasm banks is important to support breeding programs in order to identify promising accessions that can generate biofortified cultivars with high starch content in the roots. For the study of cassava accesses kept in the Active Bank of Germplasm of Eastern Amazonia, a study was carried out using quantitative descriptors and molecular marker SNP (*Single Nucleotide Polymorphism*). When evaluating 78 cassava accessions to verify the presence of SNP in a region of approximately 300 bp in the phytoene synthase 2 (*psy2*) gene, important in the synthesis of carotenoids and associated with root staining, three SNPs were identified in the coding region, and the existence of eight haplotypes. The c.1434G>C and c.1485C>A variations of SNP1 and SNP3 were found in root accesses of yellow and cream coloring respectively, and nucleotide polymorphism resulted in amino acid substitution. The effect of mutations on the structure and function of said proteins was also evaluated *in silico* by molecular dynamics. The results were compatible with the phenotype detected experimentally. It was also used a genotyping-by-sequencing platform to genotype a panel of 398 accessions, where 80,327 SNPs were identified in the genome of cassava. Within a genomic region of chromosome 1, a SNP associated with starch content was identified in cassava. This SNP was positioned in the vicinity of the Manes.01G048600 locus and functional annotation of this gene relates its biological activity to the enzyme glucose-6-phosphate 1-epimerase, which is strongly linked to carbohydrate metabolism. Therefore, these SNPs may represent an important marker for assisted selection in breeding programs aimed at increasing starch content and biofortifying

roots. In addition, the chemical characterization of 46 accessions of "bitter" and "sweet" cassava was carried out to verify the genetic relationships among accessions based on chemical characteristics. It was evaluated: moisture, ash, total soluble solids, total titratable acidity, pH, total carotenoids, free and total cyanide, crude protein, glucose, fructose, sucrose and starch. A considerable phenotypic variation was detected for the characters. In addition to free and total cyanide, characteristics such as sugars and moisture had different averages for "bitter" and "sweet" cassava. Starch and free and total cyanide concentrations were the least important characters to discriminate accesses. The "bitter" and "sweet" cassava were grouped into different clusters in the dendrogram and in the dispersion of principal component analysis. The germplasm bank of the Eastern Amazon presented genetic variation for chemical characteristics of the cassava root.

Keywords: Starch. Chemical characterization. Carotenoids. Genotyping-by-sequencing. Haplotypes. Protein Modeling. Polymorphism.

1 CONTEXTUALIZAÇÃO

A mandioca (*Manihot esculenta* Crantz) pertence à família Euphorbiaceae e é uma planta tropical originada na Amazônia brasileira, incluindo especificamente os Estados de Mato Grosso, Rondônia e Acre (OLSEN, 2004; LÉOTARD et al. 2009). Além disso, caracteriza-se por apresentar parte aérea arbustiva lenhosa e perene com altura entre 1 a 5 m dependendo do tipo de cultivar; sendo considerada tolerante às condições de seca e a baixa fertilidade do solo (FREGENE & PUONTI-KAERLAS, 2002; SAENGCHAN et al. 2009).

Botanicamente, demonstra ser uma espécie diplóide, com $2n=36$ cromossomos e meiose regular de $n=18$ (CARVALHO & GUERRA, 2002; NASSAR, 2006). Ao longo do tempo a mandioca, vem sendo propagada vegetativamente pela interferência humana, contudo manteve a reprodução sexuada ativa, importante para a variabilidade genética e possibilitando aos melhoristas selecionar acessos (genótipos) de maior importância agrônômica (SILVA et al. 2001).

Cultivada em regiões tropicais e subtropicais da Ásia, América Latina e África a produção mundial de mandioca alcançou 288,4 milhões de toneladas em 2016, enquanto a produção brasileira foi responsável por 22,3 milhões de toneladas, correspondendo aproximadamente a 8% da produção mundial. Entre as regiões do Brasil, a região Norte é a maior produtora com 8,2 milhões de toneladas, sendo o estado do Pará o maior produtor nacional produzindo 4,2 milhões de tonelada (FAO, 2016; AGRIANUAL, 2016) (FAO, 2016; AGRIANUAL, 2016).

Historicamente, essa espécie tem contribuído de forma importante para a segurança alimentar, sendo uma importante fonte de calorias para mais de 800 milhões de pessoas em todo o mundo (LEBOT, 2009; SUJA et al. 2010). Além de ser consumida na forma de alimento fresco, é rica em amido, que pode ser usado na indústria farmacêutica, têxtil, cosmético, petróleo, produtos de papel biodegradável e indústrias de celulose, ração animal e na produção de álcool (REINHARDT et al. 2013).

A mandioca é uma planta cianogênica, pois apresenta glicosídeos cianogênicos nas raízes, principalmente linamarina e lotaustralina, que podem ser hidrolisados enzimaticamente e produzir ácido cianídrico por um processo chamado cianogênese (SÁNCHEZ et al. 2009). Desse modo, BOLHUIS (1954) classificou a mandioca em dois grupos, de acordo com o potencial cianogênico: “mandiocas bravas ou amargas”, cuja concentração de glicosídeos cianogênicos está acima de 100 mg kg^{-1} , o que as tornam altamente tóxicas para o consumo humano ou animal; e as “mandiocas mansas” ou “mandioca doce ou macaxeira”, cujo nível de

compostos cianogênicos é inferior a 100 mg kg^{-1} . Dentro do grupo das “mandiocas doce ou macaxeiras” também existe um grupo denominado “mandiocaba” ou “mandioca açucarada”, esse grupo de mandioca apresenta alto teor de água e açúcares livre, como glicose (22%), manose galactose, arabinose e raminose. Porém as mandiocabas apresentam baixos teores de matéria seca e amido (2%), sendo utilizada para a produção de uma bebida conhecida como manicueira e com grande potencial para a produção de etanol (CARVALHO et al., 2004). CARVALHO et al. (2004) testaram a expressão de vários genes que codificam as enzimas relacionadas ao metabolismo dos carboidratos e identificaram uma falha na expressão do gene que codifica a enzima SBE (branching enzyme) responsável pela ramificação do amido, e podendo ser a causa da presença de altas quantidades de açúcar livre e da produção de um amido similar ao glicogênio na mandiocaba.

As variedades de mandioca bravas são utilizadas para produção de farinha, extração de amido e outros produtos, consumido após algum tipo de processamento industrial. Variedades de mandioca doce são mais versáteis, podem ser destinadas ao processamento, igual às variedades bravas, e também consumidas após preparos mais simples como cozimento, fritura ou usadas para preparar pratos típicos da região norte e nordeste (VALLE et al. 2004).

Apesar da mandioca ser um alimento consumido por milhões de pessoas, o amido é o principal nutriente de reserva da raiz (CEBALLOS et al. 2006), sendo a raiz de mandioca considerada pobre em micronutrientes como zinco, ferro e pró-vitamina A, limitando sua qualidade nutricional. Estudos vem sendo realizados para sanar problemas de deficiência nutricional em regiões onde essa cultura é a base alimentar da população. Desse modo, a mandioca foi incluída em programas de biofortificação que tem por objetivo melhorar a qualidade nutricional da raiz através do melhoramento genético. Para a obtenção de cultivares biofortificadas, é necessário fornecer aos programas de melhoramento de mandioca informações sobre variabilidade genética, por meio da caracterização e avaliação de germoplasma, para identificar genótipos que possam ser fontes de carotenóides, proteína, baixo nível de cianeto e alta produção de amido (CHAVÉZ et al. 2005; CARVALHO et al. 2011; VIEIRA et al. 2011). Entretanto, poucos estudos visam à caracterização química das raízes de mandioca tradicionalmente cultivadas na região amazônica, e sua utilização no melhoramento genético.

Nos últimos anos, trabalhos visando a biofortificação da mandioca resultaram no aumento considerável da concentração de carotenóides na raiz. CHÁVEZ et al. (2005) avaliaram 1.789 genótipos de mandioca do banco de germoplasma do Centro Internacional de Agricultura Tropical (CIAT), encontrando concentração de carotenóides totais variando de

1.02 a 10.4 $\mu\text{g g}^{-1}$ de raiz de mandioca. SÁNCHEZ et al. (2014) estudaram 3395 genótipos de mandioca e encontraram uma variação de carotenóides totais de 0.2 a 25.5 $\mu\text{g g}^{-1}$ em base úmida, incluindo materiais melhorados por cruzamentos direcionados, confirmando a eficiência dos programas de melhoramento convencional, que podem aumentar de forma gradual no valor nutricional e melhorar a biodisponibilidade de carotenóides nas raízes de mandioca (CEBALLOS et al. 2012). Esses resultados comprovam o potencial da raiz de mandioca como fonte de carotenóides, especialmente β -caroteno, precursor da vitamina A (IGLESIAS et al. 1997).

O aumento da concentração de carotenóides pró-vitamina A é importante para saúde humana, pois esse micronutriente é essencial para o funcionamento normal da visão, sistema imunológico, crescimento e desenvolvimento, manutenção da integridade celular epitelial e para a reprodução (ACC / SCN, 2000). Assim, o consumo de raiz de mandioca, mesmo em concentrações moderadas de carotenóides entre 6 a 10 $\mu\text{g g}^{-1}$ em raiz fresca, pode ajudar no combate à deficiência de vitamina A (VIMALA et al. 2011). Portanto trabalhos que visem a biofortificação da mandioca representam uma estratégia sustentável, que pode ser alcançada através de técnicas de melhoramento, os quais utilizam a alta variabilidade genética existente na espécie para desenvolver variedades de mandioca com melhor qualidade nutricional (CHAVÉZ et al. 2005; WELCH & GRAHAM, 2005).

A ampla variabilidade genética da cultura pode ser decorrente da facilidade de polinização cruzada, elevada heterozigose natural da espécie, deiscência dos frutos e as ocasionais utilização de sementes na propagação (MCKEY et al. 2010; MONTERO-ROJAS et al. 2011). Atualmente a maior parte da diversidade genética de mandioca utilizada no mundo vem do Brasil, isso se deve ao fato dessa espécie ser originada na Amazônia brasileira (LÉOTARD et al. 2009; OLIVEIRA et al. 2014).

Considerando que a mandioca é cultivada em praticamente toda a extensão territorial do Brasil e que a espécie possui ampla variabilidade genética para adaptação a diferentes ambientes e condições climáticas, o país vem mantendo essa variabilidade em banco de ativo de germoplasma (BAG) e coleções, situadas em diferentes locais. Um deles está estabelecido no Estado do Pará, principal produtor de mandioca no Brasil. As bases deste BAG são acessos amostrados em propriedades de agricultores familiares do Estado do Pará. Atualmente, esse BAG é constituído de 470 acessos, incluindo acessos de mandioca "doce", "brava" e "açucarado" (MOURA et al., 2013). Os esforços para o melhoramento da mandioca em todo o mundo resultaram em impacto significativo na produção, particularmente em termos de tolerância a doenças, produtividade e melhoria na qualidade nutricional (LEGG et al. 2006).

O uso de marcadores moleculares no melhoramento de mandioca está em fase inicial (OKOGBENIN et al. 2007), no entanto, ganhos genéticos já foram observados aplicando esse tipo de tecnologia em culturas como o arroz, milho e soja (EATHINGTON, 2005; RAGOT & LEE, 2007; SONAH et al. 2015).

Os marcadores moleculares estão no centro da genética moderna, permitindo estudos de genética de populações, ecologia, evolução e melhoramento genético, sendo pré-requisitos para trabalhos de mapeamento genético, análise de segregação, diagnóstico genético, exame forense, análise filogenética e numerosas aplicações biológicas (SEMAGN et al. 2006; SONAH et al. 2011). Entre os vários tipos de marcadores moleculares podemos citar: RAPD (amplificação arbitrária polimórfica de DNA) e RFLPs (Polimorfismo do Comprimento dos Fragmentos de Restrição), que foram utilizados pela primeira vez na mandioca para o estudo da diversidade genética do gênero *Manihot* (MARMEY et al. 1994). Mais tarde, os AFLP (polimorfismos de comprimento de fragmento amplificado) foram usados para compreender diferenciação genética em mandioca (ELIAS et al. 2000; FREGENE et al. 2000). Nos últimos anos, no entanto, estes tipos de marcadores foram amplamente substituídos por marcadores microssatélite ou SSR (Simple Sequence Repeat) e por SNP (Single Nucleotide Polymorphism- Polimorfismo de Nucleotídeo Único) (FERGUSON et al. 2012).

Portanto os marcadores moleculares, apresentam a vantagem de representarem o genótipo proveniente de um grande número de descritores agrônômicos (WUNSCH & HORMAZA, 2002). Além disso, aceleram a obtenção de genótipos desejáveis e podem estar ligados a locus que determinam características de interesse, não sendo influenciado por variações ambientais (FUGANTI et al. 2004).

Embora vários tipos de marcadores moleculares tenham sido desenvolvidos e estão sendo rotineiramente usados no melhoramento de plantas, a maior parte destes marcadores são restringidos nas suas aplicações, devido à limitação da sua disponibilidade e o custo elevado das análises efetuadas em grande escala. Entre os diversos marcadores moleculares citados, os SNP são os mais abundantes no genoma, sendo adequados em uma gama de análises de escalas genômicas (RAFALSKI, 2002; AGARWAL et al. 2008).

Os marcadores SNP tem como base as alterações elementares do DNA, ou seja, mutações em bases únicas da cadeia de bases nitrogenadas (Adenina, Citosina, Timina e Guanina). As mutações mais comuns são as transições, onde ocorrem trocas de uma purina por outra purina (A/G) ou de uma pirimidina por outra pirimidina (C/T). As mutações do tipo transversões ocorrem quando há troca de uma purina por uma pirimidina, ou vice-versa (C/T; A/G). Os marcadores SNP são bi-alélicos, sendo encontrados apenas dois variantes em uma

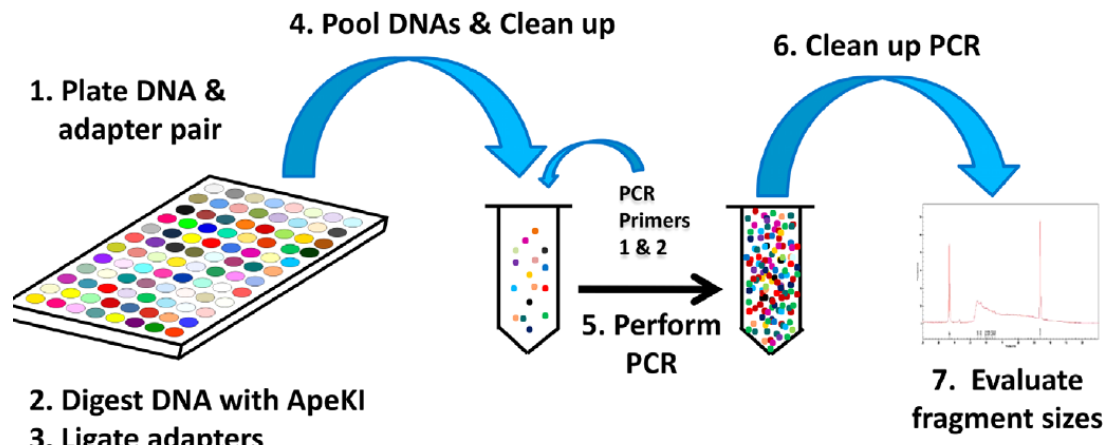
espécie (Ex: um alelo corresponde a um par de bases A/T e o outro a um G/C). Os SNPs podem ocorrer em regiões codificadoras ou com função regulatória, podendo influenciar por exemplo nas sequências de aminoácidos das proteínas e conseqüentemente na estrutura proteica e atividade enzimática, porém, na maior parte das vezes são encontrados em espaços intergênicos, sem função determinada (CAETANO, 2009).

A identificação de uma alta densidade de SNPs em mandioca facilitaria o progresso das pesquisas em genômica e de reprodução. Atualmente, foram identificados SNPs relacionados a cor da raiz de mandioca, sendo importante para a seleção de genótipos com raiz amarela, fonte de carotenóides pró-vitamina A. Assim os SNPs são uma descoberta viável para o estudo da variabilidade da mandioca, associado a características de interesse (WELSCH et al. 2010; ESUMA et al. 2016).

Os SNPs podem ser identificados por diversas plataformas de sequenciamento, como por exemplo por meio do sequenciamento de Sanger, o qual permite determinar a sequência exata de uma cadeia de DNA com até 500 nucleotídeos, sendo esse tipo de sequenciamento chamado de sequenciamento de primeira geração; No caso do sequenciamento de segunda geração, utiliza a estratégia de genotipagem por sequenciamento (*Genotyping by sequencing* (GBS)) que é um sistema *multiplex* para genotipagem do genoma, o qual permite a identificação de grande número de polimorfismos de nucleotídeo único (SNP) e pequenas inserções e deleções (*indels*) com aplicabilidade em análises genéticas (MITCHELL 2012). Assim a tecnologia GBS é uma das melhores plataformas para os estudos que vão desde a identificação de regiões polimórficas de genes individuais até perfil do genoma completo. As principais vantagens desta técnica incluem baixo custo, redução da manipulação de amostras, menos etapas de PCR e purificação, sem fracionamento do tamanho, um sistema de código de barras eficiente e facilidade para alterar a escala (DAVEY et al. 2011; BEISSINGER et al. 2013).

A técnica de GBS combina o uso de enzimas de restrições para digestão de múltiplas amostras de DNA e com isso construir as bibliotecas genômicas reduzidas; amplificação dos fragmentos resultantes e a utilização de plataformas de sequenciamento de nova geração, como o *Illumina*, na identificação de regiões polimórficas (POLAND & RIFE, 2012). Os passos para a realização da técnica são: fixação do DNA e adaptadores, digestão de DNA com enzimas de restrição, mistura dos DNA e purificação, PCR, limpeza dos produtos de PCR e análise do tamanho dos fragmentos (ELSHIRE et al. 2011), conforme a figura 1

Figura 1.1. Passos para a construção de bibliotecas de GBS



Fonte: ELSHIRE et al. (2011).

Atualmente o GBS é utilizado em diferentes espécies e é considerado uma ferramenta bastante atrativa para estudos genéticos de plantas, como por exemplo mapeamento associativo ou GWAS (*Genome Wide Association Studies*), caracterização de germoplasma, diversidade genética, genética de população, construção de mapas genéticos, identificação de QTLs, estudo da complexidade genômica, determinação de nível de ploidia e seleção genômica (DESCHAMPS et al. 2012; POLAD et al. 2012; CROSSA et al. 2013; LU et al. 2013; SONAH et al. 2013; WARD et al. 2013). A versatilidade e o potencial para uso generalizado da tecnologia GBS depende das ferramentas disponíveis para enfrentar a elevada taxa de erro e a existência de dados perdidos a partir dessa técnica. Para resolver estes problemas, vários estudos de bioinformática têm sido realizados para superar esses obstáculos (BRADBURY et al. 2007; SONAH et al. 2013).

Como mencionado anteriormente, o GBS permite a realização de estudos de associação genômica ampla (GWAS), importante no melhoramento genético de plantas, animais e também na genética humana. No das plantas, a GWAS aumenta a rapidez e eficiência no processo seletivo (GODDARD et al. 2009; CAMPOS et al. 2010; MAKOWSKY, et al. 2011).

O GWAS analisa variações na sequência do DNA de todo o genoma com a finalidade de identificar características fenotípicas de interesse. Tais, estudos de associação genômica mostram antigos eventos de recombinação para identificar locus genéticos relacionados a caracteres de interesse (DISTEFANO & TAVERNA, 2011; HAN & HAUANG, 2013). As realizações desses estudos são possíveis devido a disponibilidade da tecnologia de microarray,

baseada em chips de alta densidade (dezenas de milhares a um milhão ou mais de SNPs). Duas plataformas principais têm sido usadas para a maioria dos GWAS, *Illumina* e *Affymetrix* (DISTEFANO & TAVERNA, 2011).

GWAS fornecem uma enorme oportunidade de informações para examinar as interações entre as variantes genéticas de todo o genoma. A análise, no entanto, apresenta inúmeros desafios computacionais, estatísticos e logísticos. Devido ao fato de a maioria dos estudos envolvendo GWAS utilizarem um grande número de SNPs, examinar todas as combinações de pares de SNPs torna-se uma abordagem computacionalmente complexa, mesmo para os algoritmos altamente eficientes. Uma abordagem para solucionar este problema propõe reduzir o conjunto de SNPs genotipados, eliminando informações redundantes (BUSH & MOORE, 2012).

Com mais de 97% do genoma da mandioca sequenciado (PROCHNIK et al. 2012) é possível usar GBS e GWAS (ELSHIERE et al. 2011; OLIVEIRA et al. 2012) para ancorar com precisão os SNPs ligados a traços de interesse (RABBI et al. 2014). O uso de dados de SNP gerados a partir do GBS permitiu desenvolver mandiocas resistentes ao *geminivírus* (mosaico da mandioca). Além disso a tecnologia GBS também já foi usada para identificar SNPs associados a síntese de carotenóides, sendo importante para trabalho de melhoramento genético que visam a biofortificação da raiz de mandioca em regiões do mundo onde a cultura é a base alimentar (ESUMA et al. 2016).

Assim, o presente trabalho tem como principais objetivos: I- identificar SNP em uma região codificante do gene fitoeno sintase 2 (*psy2*), que estão associados a coloração amarela da raiz e produção de carotenoides em acessos de mandioca, originários da Amazônia brasileira e conservados em banco ativo da Amazônia Oriental; II- identificar marcadores SNPs e genes candidatos associados ao conteúdo de amido, para auxiliar o melhoramento genético sistemático da mandioca para níveis crescentes de amido nas raízes e III- avaliar caracteres químicos de raízes "bravas" e "doces" do banco de germoplasma da Amazônia Oriental.

1.1 REFERÊNCIAS

ACC/SCN. 2000. Fourth report on the world nutrition situation. Administrative Committee on Coordination, Subcommittee on Nutrition. Food Policy Research Institute, United Nations, Geneva, Switzerland.

AGARWAL, M., SHRIVASTAVA, N., AND PADH, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. **Plant Cell Rep.** v. 27, 617–631, 2008.

AGRIANUAL. **Anuário da Agricultura Brasileira. Mandioca.** São Paulo: FNP Consultoria & Comércio, 2016.

BEISSINGER, T. M.; HIRSCH, C. N.; SEKHON, R. S.; FOERSTER, J. M.; JOHNSON, J. M.; MUTTONI, G.; VAILLANCOURT, B.; BUELL, C. R.; KAEPLER, S. M.; DE LEON, N. Marker density and read depth for genotyping populations using genotyping-by-sequencing. **Genetics**, v. 193, n. 4, p. 1073-81, 2013.

BOLHUIS, G. G. The toxicity of cassava roots. **Netherlands Journal of Agricultural Science**, v. 2, p.176-185, 1954.

BRADBURY, P. J; ZHANG, Z.; KROON, D. E.; CASSTEVENS, T. M.; RAMDOSS, Y.; BUCKLER, E. S. TASSEL: Software for association mapping of complex traits in diverse samples. **Bioinformatics** 23:2633-2635, 2007.

BUSH, W. S.; MOORE, J. H. Chapter 11: Genome-Wide Association Studies. **PLoS Computational Biology**, v. 8, n. 12, p. e1002822, 27, 2012.

CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira Zootecnia.**, v.38, p. 64-71, 2009.

CAMPOS, G. de los; GIANOLA, D.; ALLISON, D. B.; Predicting genetic predisposition in humans: the promise of whole-genome markers. **Nature Reviews Genetics**, London, v. 11, p. 880-886 Dec. 2010.

CARVALHO L. J. C. B.; BATISTA DE S. C. R.; MATTOS C. J .C.; BLOCH C.; CAMPOS L. Identification and characterization of a novel cassava (*Manihot esculenta* Crantz) clone with high free sugar content and novel starch. **Plant Molecular Biology**, v. 56, n. 4, p. 643–659, 2004.

CARVALHO, L. M. J.; GOMES, P. B.; GODOY, R. L. O.; PACHECO, S.; MONTE, P. H. F.; CARVALHO, J. L. V.; NUTTI, M. R.; NEVES, A. C. L.; VIEIRA, A. C. R. A.; RAMOS, S. R. R. Total carotenoid content, α -carotene and β -carotene, of landrace pumpkins (*Cucurbita moschata* Duch): A preliminary study. **Food Research International**, 2011.

CARVALHO, R.; GUERRA, M. Cytogenetics of *Manihot esculenta* Crantz (cassava) and eight related species. **Hereditas**, v. 136, p. 159-168, 2002.

CEBALLOS H.; SANCHEZ, T.; MORANTE, N.; FREGENE, M.; DUFOUR D.; SMITH, A.; DENYER, K.; PEREZ, J.; CALLE, F.; MESTRES. C. Discovery of an Amylose-free Starch Mutant in Cassava (*Manihot esculenta* Crantz). *Journal of Agricultural and Food Chemistry*, v. 55, p. 7469 -7476, 2006.

CEBALLOS, H.; LUNA, J.; ESCOBAR, A. F.; PÉREZ, J. C.; ORTIZ, D.; SÁNCHEZ, T.; PACHÓN, H.; DUFOUR, D. Spatial distribution of dry matter in yellow fleshed cassava roots

and its influence on carotenoids retention upon boiling. **Food Research International**, v. 45, p. 52-59, 2012.

CHÁVEZ, A. L.; SÁNCHEZ, T.; JARAMILLO, G.; BEDOYA, J. M.; ECHEVERRY, J.; BOLAÑOS, E. A.; CEBALLOS, H.; IGLESIAS, C. A. Variation of quality traits in cassava roots evaluated in landraces and improved clones. **Euphytica**, v. 143, p. 125-133, 2005.

CHING, A.; CALDWELL, K.; JUNG, M.; DOLAN, M. S. O.; TINGEY, S.; MORGANTE, M. RAFALSKI, A. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. **BMC Genetcs**, v 3, p. 1-114, 2002.

CHOI, I. Y.; HYTEN, D. L.; MATUKUMALLI, L. K.; SONG, Q.; CHAKY, J. M.; QUIGLEY, C. V.; CHASE, K.; LARK, G. K.; REITER, R. S.; YOON, MUN-SUP; HWANG, E. Y.; YI, S. I.; YOUNG, N. D.; SHOEMAKER, R. C.; TASSELL, C. P. VAN; SPECHT, J. E; CREGAN, P. B. A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. **Genetics**, v.176, p. 685-96, 2007.

CROSSA, J.; BEYENE, Y.; KASSA, S.; PÉREZ, P.; HICKEY, J. M.; CHEN, C.; CAMPOS, G.; BURGUEÑO, J.; WINDHAUSEN, V. S.; BUCKLER, E; JANNINK, J.; CRUZ, M. A. L.; BABU, R. Genomic Prediction in Maize Breeding Populations with Genotyping-by-Sequencing. **Genes|Genomes|Genetics**, v. 3, n. 11, p. 1903–1926, 2013.

DAVEY, J. W.; HOHENLOHE, P. A.; ETTER, P. D.; BOONE, J. Q.; CATCHEN, J. M.; BLAXTE, M. L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics** v. 12: 499–510, 2011.

DESCHAMPS, S.; LLACA, V.; MAY, G. D. Genotyping-by-Sequencing in Plants. **Biology**, v. 1, n. 3, p. 460–483, 2012.

DISTEFANO, J. K.; TAVERNA, D. M. Technological Issues and Experimental Design of Gene Association Studies. **Methods in Molecular Biology** (Clifton, N.J.), v. 700, p. 3–16, 2011.

EATHINGTON, S. R. (2005). Practical applications in molecular technology in the development of commercial maize hybrids. Proc. 60th Ann. Corn and Sorghum Seed Res. Conf. Washington DC, USA. American Seed Trade Association.

ELIAS, M.; PANAUD, O.; ROBERT, T. Assessment of genetic variability in a traditional cassava (*Manihot esculenta* Crantz) farming system, using AFLP markers. **Heredity**, v. 85, p. 219-230, 2000.

ELSHIRE, R. J.; GLAUBITZ, J. C.; SUN, Q.; POLAND, J. A.; KAWAMOTO, K.; BUCKLER, E. S.; MITCHELL, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PloS One** 6: e19379, 2011.

ESUMA, W.; HERSELMAN, L.; LABUSCHAGNE, M. T.; RAMU, LU, F.; BAGUMA, Y.; BUCKLER, E. S.; KAWUKI, R. S. Genome-wide association mapping of provitamin A carotenoid content in cassava. **Euphytica**, v. 212, n. 1, p. 97–110, 2016.

FAO. Milk and Milk Products. **Food Outlook**, n. October, p. 54–58, 2016.

FARALDO, M. I. F.; SILVA, R. M.; ANDO, A.; MARTINS, P. S. Genetic variability of landraces of cassava in geographical regions of Brazil. **Scientia Agricola**, v.57, p.499-505, 2000.

FERGUSON, M.; RABBI, I.; KIM, D.; GEDIL, M.; LOPEZ-LAVALLE, L. A. B.; OKOGBENIN, E. Molecular Markers and Their Application to Cassava Breeding: Past, Present and Future. **Tropical Plant Biology**, v. 5, p. 95-109, 2012.

FREGENE, M.; PUONTI-KAERLAS, J. Cassava biotechnology. **In Cassava: Biology, Production and Utilization**. eds Hillocks R. J., Thresh J. M., Bellotti A. C., editors. (Wallingford: CAB International), p. 179-207, 2002.

FREGENE, M.; BERNAL, A.; DUQUE, M.; DIXON, A.; TOHME, J. AFLP analysis of African cassava (*Manihot esculenta* Crantz.) germplasm resistant to the cassava mosaic disease (CMD). **Theoretical and Applied Genetics**, v. 100, p. 678-685, 2000.

FUGANTI, R.; BENEVENTI, M. A.; SILVA, J. F. V.; ARRABALARIAS, C. A.; MARIN, S. R. R.; BINNECK, E.; NEPOMUCENO, A. L. Identificação de marcadores moleculares de microssatélites para seleção de genótipos de soja resistentes a *Meloidogyne javanica*. **Nematologia Brasileira**, v. 28, n. 02, p. 125-130, 2004.

GODDARD, M.; WRAY, N. R.; VERBYLA, K.; VISSCHIER, P. Estimating effects and making predictions from genome-wide marker data. **Statistical Science**, Hayward, v. 24, p. 517-529, 2009.

GOMEZ-ROLDAN, V.; FERMAS, S.; BREWER, P. B.; PUECH-PAGÈS, V.; DUN, E. A.; PILLOT, J. P.; LETSSE, F.; MATUSOVA, R. DANOUN, S.; PORTAIS, J. C.; PUECH-PAGÈS, H.; BÉCARD, G.; BEVERIDGE, C. A.; RAMEAU, C.; ROCHANGE, S. F. Strigolactone inhibition of shoot branching. **Nature**, v. 455, p. 189-194, 2008.

GRIVET, L.; GLASZMANN, J. C.; VINCENTZ, M.; SILVA, F.; ARRUDA, P. ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. **Theoretical and Applied Genetics**, v. 106, p. 190-197, 2003.

HAN, B.; HUANG, X. Sequencing-based genome-wide association study in rice. **Current Opinion in Plant Biology**, v. 16, n. 2, p. 133-138, 2013.

IGLESIAS, C.; MAYER, J.; CHAVEZ, L.; CALLE, F. Genetic potential and stability of carotene content in cassava roots. **Euphytica**, v. 94, p. 367-373, 1997.

KAWUKI, R.; FERGUSON, M.; LABUSCHAGNE, M.; HERSELMAN, L.; KIM, D. J. Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). **Molecular Breeding**, v. 23, p. 669-684, 2009.

LEBOT, V. **Tropical root and tuber crops: cassava, sweet potato, yams and aroids**. Crop Production Science in Horticulture. CABI, Wallingford, UK. 2009.

LEGG, J. P.; B. OWOR, B.; SSERUWAGI, P.; NDUNGURU, J. Cassava mosaic virus disease in east and central africa: epidemiology and management of a regional pandemic. **Advances in virus research**, v. 67, 355-418, 2006.

LÉOTARD, G.; DUPUTIE, KJELLBERG, F.; DOUZERY, E.J.P.; DEBAIN, C.; GRANVILLE, J. J.; MCKEY, D. Phylogeography and the origin of cassava: new insights from northern rim of the Amazonia basin. **Molecular Phylogenetics and Evolution**, v. 53, p. 329-334, 2009.

LOPEZ, C.; PIEGU, B.; COOKE, R.; DELSENY, M.; TOHME, J.; VERDIER, V. Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihot esculenta* Crantz). **Theoretical and Applied Genetics**, v. 110, p. 425-431, 2005.

LU, F.; LIPKA, A.E.; GLAUBITZ, J.; ELSHIRE, R.; CHERNEY, J.H.; CASLER, M.D.; BUCKLER, E.S.; COSTICH, D.E. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. **PLoS Genet.** v. 9, n. 1, e1003215, 2013.

MAKOWSKY, R.; PAJEWSKI, N. M.; KLIMENTIDIS, Y. C.; VAZQUEZ, A. I.; DUARTE, C. W.; ALLISON, D. B.; CAMPOS, G. de los. Beyond missing heritability: prediction of complex traits. **Plos Genetics**, v. 7, n. 4, p. 1-15 2011.

MARMEY, P.; BEECHING, J.; HAMON, S.; CHARRIER, A. Evaluation of cassava (*Manihot esculenta* Crantz.) germplasm using RAPD markers. **Euphytica**, v.74, p.203-209, 1994.

MCKEY, D.; ELIAS, M.; PUJOL, B.; DUPUTIE, A. The evolutionary ecology of clonally propagated domesticated plants. **New Phytologist**, v.186, p. 318-332, 2010.

MITCHELL, S. 2012. Introduction to Genotyping by Sequencing (GBS Workshop). Disponível em: <http://mediasite.video.cornell.edu/mediasite/Catalog/pages/catalog.aspx?catalogId=a0f3a0d9-1523-4c3a-8729-2b8cfc1f9160>. Acesso em: 20/04/17.

MONTERO-ROJAS, M.; CORREA, A. M.; SIRITUNGA, D. Molecular differentiation and diversity of cassava (*Manihot esculenta*) taken from 162 locations across Puerto Rico and assessed with microsatellite markers. **AoB Plants**, p. 1-13, 2011.

MOURA, E. F.; FARIAS NETO, J. T; SAMPAIO, J. E.; SILVA, D. T.; RAMALHO, G. F. Identification of duplicates of cassava accessions sampled on the North Region of Brazil using microsatellite markers. **Acta Amazonica**, v. 43, n. 1, p. 461–468, 2013.

NASSAR, N. M. Cassava genetic resources: extinct everywhere in Brazil. **Genetic Resources and Crop Evolution**, v. 53, p. 975-983, 2006.

OKOGBENIN, E.; PORTO, M.; EGESI, C.; MBA, C.; ESPINOSA, E.; SANTOS, L.; OSPINA, C.; MARIN, J.; BARRERA, E.; GUTIERREZ, J.; EKANAYAKE, C.; IGLESIAS, C.; FREGENE, M. A. Marker-assisted introgression of resistance to cassava mosaic disease into Latin American germplasm for the genetic improvement of cassava in Africa. **Crop Science**, v. 47, p. 1895-1904, 2007.

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**, v. 127, n. 6, p. 1423–1440, 2014.

OLIVEIRA, E. J.; RESENDE, M. D. V.; SILVA SANTOS, V et al. Genome-wide selection in cassava. **Euphytica**, v.187, p. 263–276, 2012.

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

OLSEN, K. M.; SCHAAL, B. A. Microsatellite variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a Southern Amazonian origin of domestication. **American Journal of Botany**, v. 88, p. 131-142, 2001.

POLAND, J. A.; RIFE, T. W. Genotyping-by-Sequencing for Plant Breeding and Genetics. **The Plant Genome**, v. 5, n. 3, 2012.

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

QUELLER, D.; STRASSMAN J.; HUGHES C. Microsatellites and Kinship. **Trends in Ecology Evolution.**, v. 8, p. 285-288, 1993.

RABBI, I. Y.; HAMBLIN, M. T.; KUMAR, P. L.; GEDIL, M. A.; IKPAN, A. S.; JANNINK, J. L.; KULAKOW, P. A. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. **Virus Res**, v.186, p.87–96, 2014.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**, v. 5, p. 94–100, 2002.

RAGOT, M.; LEE, M. (2007). Marker-assisted selection in maize: current status, potential limitation and perspectives from the private and public sectors. Marker assisted selection: current status and future perspectives in crops, livestock, forestry and fish. FAO, Rome, pp 117-150.

REINHARDT, H.; NEBAMBI, L.; GRAEME, T. Save and GrowCassava: A Guide to Sustainable Production Intensification, first ed. Food and Agriculture Organization of the United Nations (FAO), Rome. 2013.

SAENGCHAN, K.; NOPHARATANA, A.; SONGKASIRI, W. Enhancement of tapioca starch separation with ahydrocyclone: effects of apex diameter, feed concentration, and pressure drop on tapioca starch separation with ahydrocyclone. **Chem. Eng. Process.: Process Intensification**, v. 48, p.195-202, 2009.

SÁNCHEZ, T.; CEBALLOS, H.; DUFOUR, D; ORTIZ, D.; MORANTE, N.; CALLE, F.; FELDE, T.Z.; DOMÍNGUEZ, M.; DAVRIEUX, F. Prediction of carotenoids, cyanide and

dry matter contents in fresh cassava root using NIRS and Hunter color techniques. **Food Chemistry**, v. 151, p. 444-451, 2014.

SÁNCHEZ, T.; SALCEDO, E.; CEBALLOS, H.; DUFOUR, D.; MAFLA, G., MORANTE, N.; CALLE, F.; PÉREZ, J. C.; DEBOUCK, D.; JARAMILLO, G.; MORENO, I. X. Screening of starch quality traits in cassava (*Manihot esculenta* Crantz). **Starch/Stärke**, v. 61, p. 12-19, 2009.

SCHWARTZ, S. H.; TAN, B. C.; GAGE, D. A.; ZEEVAART, J. A. D; MCCARTY, D. R. Specific oxidative cleavage of carotenoids by VP14 of maize. **Science**, v. 276, p. 1872-1874, 1997.

SEMAGN, K.; BJORNSTAD, A.; NDJIONDJOP, M. N. Principles, requirements and prospects of genetic mapping in plants. **African Journal of Biotechnology**, v. 5 (25), p. 2569-2587, 2006.

SILVA, R. M.; BANDEL, G.; FARALDO, M. I. F.; MARTINS, P. S. Biologia reprodutiva de etnovarietades de mandioca. **Scientia agricola**, v. 58, p. 101-107, mar., 2001.

SONAH, H.; BASTIEN, M.; IQUIRA, E.; TARDIVEL, A.; LEGARE, G. et al. An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. **PLoS One**, v. 8, n.1, 2013.

SONAH, H.; DESHMUKH, R. K.; SHARMA, A.; SINGH, V. P.; GUPTA, D. K.; GACCHE, R. N.; RANA, J. C.; SINGH, N. K.; SHARMA, T. R. Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. **PLoS ONE**, v. 6, p. 291– 298, 2011.

SONAH, H.; O'DONOUGHUE, L.; COBER, E.; RAJCAN, I.; BELZILE, F. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. **Plant Biotechnology Journal**, v. 13, n. 2, p. 211–221, 2015.

SUJA, G.; JOHN, K.S.; SREEKUMAR, J.; SRINIVAS, T. Short-duration cassava genotypes for crop diversification in the humid tropics: growth dynamics, biomass, yield and quality. **Journal of the Science of Food and Agriculture**, v. 90, p. 188-198, 2010.

UMEHARA, M.; HANADA, A.; YOSHIDA, S.; AKIYAMA, K.; ARITE, T.; TAKEDA-KAMIYA, N.; MAGOME, H.; KAMIYA, Y.; SHIRASU, K.; YONEYAMA, K.; KYOZUKA, J.; YAMAGUCHI, S. Inhibition of shoot branching by new terpenoid plant hormones. **Nature**, v. 455, p. 195-200, 2008.

VALLE, T. L.; CARVALHO, C. R. L.; RAMOS, M. T. B.; MÜHLEN, G. S.; VILLELA, O. V. Conteúdo cianogênico em progênies de mandioca originadas do cruzamento de variedades mansas e bravas. **Bragantia**, v.63, p.221-226, 2004.

VIEIRA, E. A.; FIALHO, J. F.; FALEIRO, F. G.; BELLON, G.; SILVA, M. S. Caracterização molecular de acessos de mandioca biofortificados com potencial de uso no melhoramento genético. **Revista Ciência Agronômica** 42: 457-463, 2011.

VIMALA, B.; THUSHARA, R.; NAMBISAN, B; SREEKUMAR, J. Effect of processing on the retention of carotenoids in yellow-fleshed cassava (*Manihot esculenta* Crantz) roots. **International Journal of Food Science and Technology**, v. 46, p. 166-169, 2011.

WARD, J. A.; BHANGOO, J.; FERNÁNDEZ-FERNÁNDEZ, F.; MOORE, P.; SWANSON, J. D.; VIOLA, R.; VELASCO, R.; BASSIL, N.; WEBER, C. A.; SARGENT, D. J. Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. **BMC Genomics**, 2-14, 2013.

WELCH, R. M.; GRAHAM R. D. Agriculture: The real nexus for enhancing bioavailable micronutrients in food crops. **Journal of Trace Elements in Medicine and Biology**., v. 18, p. 299-307, 2005.

WELSCH, R.; ARANGO, J.; BÄR, C.; SALAZAR, B.; AL-BABILI, S.; BELTRÁN, J.; CHAVARRIAGA, P.; CEBALLOS, H.; TOHME, J.; BEYER, P. Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. **The Plant Cell**, v. 22, p. 3348-3356, 2010.

WUNSCH, A.; HORMAZA, J. I. Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. **Euphytica**, v. 125, p. 59-67, 2002.

ZHU, Y. L.; SONG, Q. J.; HYTEN, D.L.; VAN TASSELL, C.P.; MATUKUMALLI, L. K.; GRIMM, D. R.; HYATT, S. M.; FICKUS, E. W.; YOUNG, N. D. CREGAN, P. B. Single-nucleotide polymorphisms in soybean. **Genetics**, v. 163, p. 1123-1134, 2003.

2 IDENTIFICAÇÃO DE SNP NO GENE FITOENO SINTASE 2 (*psy2*) ASSOCIADOS A COLORAÇÃO AMARELA DA RAIZ EM ACESSOS DE MANDIOCA DA AMAZÔNIA BRASILEIRA¹

RESUMO

A mandioca (*Manihot esculenta*) é uma cultura importante na alimentação humana. O presente estudo analisou 78 acessos de mandioca da Amazônia brasileira para verificar a presença de marcadores moleculares do tipo *Single Nucleotide Polymorphism* (SNP) no gene fitoeno sintase 2 (*psy2*) associados a coloração da raiz. A partir do sequenciamento de uma região de aproximadamente 300 pb contendo parte do íntron 2 e o éxon 3 do gene *psy2*, foram identificados três SNPs na região codificante, e a existência de oito haplótipos. As variações c.1434G>C e c.1485C>A dos SNP1 e SNP3 foram encontrados em acessos de raiz de coloração amarela e creme e o polimorfismo de nucleotídeo resultou na substituição de aminoácidos. O SNP2 foi identificado em acessos com raízes brancas e suas variações alélicas não foram associadas a cor da raiz. Quatro haplótipos que apresentaram o SNP3 foram amplamente distribuídos nos Estados do Pará, Amapá, Rondônia e Amazonas. Todos os acessos de raiz amarela ou creme que apresentaram o SNP3 estavam em heterozigose, confirmando a dominância do SNP, exceto o genótipo CPATU 018Y. As variações do gene *psy2* que apresentaram o SNP3, com substituição do aminoácido alanina por ácido aspártico foram predominantes em acessos de raiz amarela. O efeito das mutações na estrutura e função das proteínas também foi avaliado *in silico* por dinâmica molecular. Os resultados foram compatíveis com o fenótipo do organismo, detectado experimentalmente. Portanto, o SNP3 pode representar um importante marcador para seleção assistida em programas de melhoramento genético visando biofortificação de raízes.

Palavras-chave: Carotenóides. Haplótipo. Melhoramento genético. Modelagem de proteína. Polimorfismo.

¹ Este capítulo segue as normas da Revista Plant Molecular Biology

ABSTRACT

Cassava (*Manihot esculenta*) is an important food crop. The present study analyzed 78 accessions of cassava from the Brazilian Amazon to verify the presence of *single nucleotide polymorphism* (SNP) molecular markers in the phytoene synthase 2 (*psy2*) gene associated with root staining. From the sequencing of a region of approximately 300 bp containing part of the intron 2 and the exon 3 of the *psy2* gene, three SNPs located in the coding region, and the existence of eight haplotypes were identified. The c.1434G>C and c.1485C>A variations of SNP1 and SNP3 were found in yellow and cream root accesses and nucleotide polymorphism resulted in amino acid substitution. SNP2 was identified in accessions with white roots and their allelic variations were not associated with root color. Four haplotypes that presented SNP3 were widely distributed in the states of Pará, Amapá, Rondônia and Amazonas. All accesses of yellow or cream root that presented the SNP3 were in heterozygosity, confirming the dominance of SNP, except the genotype CPATU 018Y. The variations of the *psy2* gene that presented the SNP3, with substitution of the amino acid alanine by aspartic acid were predominant in yellow root accesses. The effect of mutations on the structure and function of proteins was also evaluated *in silico* by molecular dynamics. The results were compatible with the phenotype detected experimentally. Therefore, this SNP can represent an important marker for assisted selection in breeding programs for root biofortification.

Keywords: Carotenoids. Genetical enhancement. Haplotypes. Polymorphisms. Protein Modeling

2.1 Introdução

A mandioca (*Manihot esculenta* Crantz) é classificada como a terceira maior fonte de carboidratos do mundo, consumida por mais de 800 milhões de pessoas, e cultivada em países da África tropical, Ásia e América do Sul. O Brasil que é considerado o centro de origem e é onde atualmente está a maior parte da diversidade genética da mandioca (Howeler et al. 2013; Niba et al. 2002). Além de ser fonte de carboidratos, acessos de mandioca podem ser uma importante fonte de carotenóides (Sánchez et al. 2014; Nassar et al. 2009).

Os carotenóides são compostos isoprenóides sintetizados nos plastídios de células vegetais e desenvolvem funções antioxidantes, atuando como pigmentos acessórios na fotossíntese, precursores na biossíntese de ácido abscísico e substratos para produção de

compostos voláteis em flores, frutas e sementes (Moise et al. 2014; Kermode 2005). A função mais importante dos carotenóides em animais é como fonte de pró-vitamina A. Carotenóides pró-vitamina A, a exemplo do β -caroteno, podem ser convertidos por clivagem oxidativa em vitamina A (retinol) (Combs 2012; Stahl e Sies 2005). Deficiência de vitamina A pode resultar em cegueira noturna e aumento da susceptibilidade a infecções podendo eventualmente resultar em morte (Combs 2012).

A concentração de carotenóides em tecidos não fotossintetizantes é limitada pela expressão do gene fitoeno sintase (*psy*) (Maass et al. 2009). A mandioca apresenta três genes *psy*, conhecidos por *psy1*, *psy2* e *psy3* (Arango et al. 2010). O gene *psy2* foi identificado como o principal envolvido na via biossintética de carotenóides, especialmente do β -caroteno, precursor de vitamina A, principalmente devido à polimorfismo de nucleotídeo único (SNP) relacionados a biossíntese de carotenóides em raiz de mandioca (Owens et al. 2014; Welsch et al. 2010).

A atividade da proteína PSY na via biossintética de carotenóides foi estudada em diversas culturas, a exemplo do milho e arroz (Shumskaya et al. 2012); em tomate, onde foi estudado mutações pontuais no gene fitoeno sintase (*psy*) que causa diferenças no teor de carotenóides durante o amadurecimento (Gady et al. 2012), em cenoura e em canola (Arango et al. 2014; López-Emparán et al. 2014). Marcadores moleculares SNP (*Single Nucleotide Polymorphisms*) são variações alélicas de base única e podem ser identificados em regiões codificantes e não codificantes do genoma (Brookes 1999). Em mandioca, ao avaliar quatro variedades com diferentes tonalidades e concentração de carotenoides, foi verificado que a presença de um polimorfismo do tipo SNP no gene fitoeno sintase 2 (*psy2*) foi capaz de produzir aproximado duas vezes mais fitoeno, indicando que um aumento da atividade enzimática (Welsch et al. 2010). Harjes et al. (2008) estudando a cultura do milho, encontraram polimorfismos no gene licopeno ϵ -ciclase (*LCYE*), assim como Yan et al. (2010) identificaram polimorfismo no gene β -caroteno hidroxilase ambos capazes de aumentar a síntese de carotenóides na espécie. Em melancia amarelas e vermelhas foram identificados SNPs no gene licopeno β -cicloase (*LCYB*), também importante para a síntese de carotenóides pró-vitamina A no fruto (Bang et al. 2007).

O estudo de genes envolvidos na síntese de carotenóides em diversas culturas pode ter impactos positivos na agricultura (Owens et al. 2014), principalmente em regiões pobres, de modo que a detecção de marcadores moleculares SNP associados a esses genes podem ser uma ferramenta útil para a seleção de acessos com características desejáveis em programas de melhoramento genético da mandioca, e com isso desenvolver variedades locais com alta

concentração de carotenóides que possam combater a deficiência de vitamina A. O objetivo desse trabalho foi identificar SNPs em uma região codificante do gene fitoenol sintase 2 (*psy2*), verificar a sua associação com a coloração da raiz de acessos de mandioca e avaliar de que forma esses SNPs afetam a estrutura da proteína PSY2 importante para a síntese de carotenóides.

2.2 Material e métodos

2.2.1 Material vegetal

Foram utilizados 78 acessos de mandioca, sendo 29 macaxeira e 49 mandioca brava, pertencentes ao BAG da Embrapa Amazônia Oriental, localizado em Belém, Pará, Brasil. Os acessos foram coletados em diversas cidades dos Estados do Pará, Amazonas, Amapá e Rondônia, pertencentes a região Amazônica (Tabela 2.1) e apresentaram cor da polpa de raiz branca (W), creme (C) e amarela (Y). A origem geográfica de nove acessos é desconhecida.

Os diferentes padrões na cor da raiz foram avaliados pelo Teste Exato de Fisher (dois lados) usando o software R (R Core Team 2013).

Tabela 2.1. Informações dos acessos de mandioca (*Manihot esculenta*, Crantz) pertencentes ao Banco Ativo de Germoplasma (BAG) da Amazônia Oriental, localizado em Belém, Pará, Brasil.

Número	Códigos	Cidade	Estado	Cor
1	CPATU 396	Acará	Pará	Branca
2	CPATU 390	Acará	Pará	Branca
3	CPATU 463	Acará	Pará	Amarela
4	CPATU 057	Alenquer	Pará	Branca
5	CPATU 437	Belém	Pará	Creme
6	CPATU 217	Belém	Pará	Branca
7	CPATU 387	Belém	Pará	Branca
8	CPATU 366	Belém	Pará	Branca
9	CPATU 317	Belterra	Pará	Amarela
10	CPATU 328	Belterra	Pará	Amarela
11	CPATU 01	Bragança	Pará	Branca
12	CPATU 160	Bragança	Pará	Amarela
13	CPATU 101	Calama	Rondônia	Amarela
14	CPATU 358	Castanhal	Pará	Branca
15	CPATU 402	Castanhal	Pará	Branca
16	CPATU 357	Castanhal	Pará	Branca

17	CPATU 462	Curuçá	Pará	Amarela
18	CPATU 190	Ferreira Gomes	Amapá	Amarela
19	CPATU 177	Ferreira Gomes	Amapá	Branca
20	CPATU 206	Ferreira Gomes	Amapá	Amarela
21	CPATU 216	Igarapé Açú	Pará	Branca
22	CPATU 498	Igarapé Açú	Pará	Branca
23	CPATU 480	Inhagapi	Pará	Branca
24	CPATU 034	Marabá	Pará	Branca
25	CPATU 454	Marapanim	Pará	Branca
26	CPATU 455	Marapanim	Pará	Branca
27	CPATU 452	Marapanim	Pará	Branca
28	CPATU 159	Medicilândia,	Pará	Amarela
29	CPATU 424	Melgaço	Pará	Amarela
30	CPATU 373	Moju	Pará	Branca
31	CPATU 376	Moju	Pará	Branca
32	CPATU 229	Nova Timboteua	Pará	Amarela
33	CPATU 117	Oiapoque	Amapá	Amarela
34	CPATU 118	Oiapoque	Amapá	Amarela
35	CPATU 272	Paragominas	Pará	Branca
36	CPATU 285	Paragominas	Pará	Amarela
37	CPATU 266	Paragominas	Pará	Amarela
38	CPATU 507	Paragominas	Pará	Amarela
39	CPATU 181	Pedra Branca do Amapari	Amapá	Branca
40	CPATU 180	Pedra Branca do Amapari	Amapá	Branca
41	CPATU 167	Pedra Branca do Amapari	Amapá	Amarela
42	CPATU 010	Portel	Pará	Amarela
43	CPATU 173	Porto Grande	Amapá	Amarela
44	CPATU 169	Porto Grande	Amapá	Creme
45	CPATU 05	Santa Luzia do Pará	Pará	Amarela
46	CPATU 238	Santa Luzia do Pará	Pará	Branca
47	CPATU 306	Santarém	Pará	Branca
48	CPATU 344	Santarém	Pará	Branca
49	CPATU 309	Santarém	Pará	Amarela
50	CPATU 315	Santarém	Pará	Amarela
51	CPATU 033	Santarém	Pará	Amarela
52	CPATU 330	Santarém	Pará	Amarela
53	CPATU 333	Santarém	Pará	Amarela
54	CPATU 018	Santarém	Pará	Amarela
55	CPATU 346	Santarém	Pará	Amarela
56	CPATU 327	Santarém	Pará	Branca
57	CPATU 02	Santarém	Pará	Branca
58	CPATU 304	Santarém	Pará	Creme

59	CPATU 352	Santarém	Pará	Branca
60	CPATU 340	Santarém	Pará	Branca
61	CPATU 308	Santarém	Pará	Branca
62	CPATU 310	Santarém	Pará	Branca
63	CPATU 321	Santarém	Pará	Branca
64	CPATU 128	São Domingos do Capim	Pará	Branca
65	CPATU 061	São Gabriel da Cachoeira	Amazonas	Amarela
66	CPATU 140	Tabatinga	Amazonas	Amarela
67	CPATU 138	Tabatinga	Amazonas	Amarela
68	CPATU 516	Tomé Açu	Pará	Amarela
69	CPATU 147	Viseu	Pará	Branca
70	CPATU 023	-	-	Branca
71	CPATU 162	-	-	Branca
72	CPATU 025	-	-	Branca
73	CPATU 021	-	-	Branca
74	CPATU 009	-	Pará	Branca
75	CPATU 026	-	Amazonas	Amarela
76	CPATU 03	-	-	Amarela
77	CPATU 036	-	-	Amarela
78	CPATU 04	-	-	Creme

2.2.2 Extração e quantificação do DNA total

O DNA foi extraído de acordo com a metodologia de Doyle e Doyle (1990) com modificações. Aproximadamente 1 g de folhas jovens foram coletadas, lavadas e maceradas com nitrogênio líquido e polivinilpirrolidona (PVP). Em seguida, 3 mL de tampão de extração CTAB (2% de CTAB, 5 M de NaCl, 0,5 M de EDTA, PVP, 1M de TrisHCl e água destilada) foi adicionado a maceração. As amostras foram homogeneizadas e incubadas a 65°C por uma hora. Depois desse período, foi adicionado ao extrato clorofórmio: álcool isoamílico (24:1). O extrato foi homogeneizado e as amostras foram centrifugadas por 10 min a 14.869 x g. O precipitado foi decantado e adicionou-se 3 mL de álcool etílico a 95% ao sobrenadante, visando precipitar o DNA. As amostras foram novamente centrifugadas durante 10 minutos a 14.869 x g. Após essa fase, o precipitado foi lavado com álcool etílico a 70% e centrifugado durante 10 minutos a 7.434 x g. As amostras de DNA foram ressuspensas com 300 µL de tampão TE (10 mM Tris-HCl, 1 mM EDTA, pH 8,0) e RNase, e quantificadas em gel de agarose a 1% e corados com brometo de etídio para comparação com o DNA do fago lambda nas concentrações de 50, 100 e 200 ng ul⁻¹. Os perfis dos géis foram visualizados em

transiluminador sob luz ultravioleta, fotodocumentados e utilizados na avaliação. Em seguida, o DNA foi diluído para 50 ng μl^{-1} e armazenado em freezer – 80 °C.

2.2.3 Reação em cadeia da polimerase (PCR)

Os primers utilizados na realização da PCR foram o *forward* (F): 5'TCTGCAAGAATGAGATTGGGTA3', e *reverse* (R): 5'GGCTAACCTGAATATCAACAGGA3', e foram desenhados para amplificar uma região de aproximadamente 300 pb do gene *psy2*, que segundo Welsch et al. (2010) apresentaram SNPs associados com a cor da raiz de mandioca (Figura 2.1).

As PCR foram preparadas para um volume final de 50 μL , contendo 50 ng de DNA total, 5 U/ μl de Taq DNA polimerase Platinum da Invitrogen, 25,4 μL de água mili-Q (ultra pura), 1X de Tampão (50 mM KCl, 10 mM Tris-HCl, pH 8,8, 0,1% de Triton X-100; 1,5 mM MgCl_2), 50mM de MgCl_2 , 1mM de dNTPs, 10 mg/mL de BSA, 20 nmol dos primers *forward* e *reverse*.

O programa de amplificação utilizado teve fase inicial de desnaturação a 95 °C por 5 min, seguida de 35 ciclos nas seguintes condições: 1 min para desnaturação a 92 °C; temperatura de anelamento padronizada em 58 °C por 1 min; 1 min para extensão dos primers a 72 °C. A extensão final foi realizada a 72 °C por 10 min. As PCR's foram realizadas em termociclador (Applied Biosystems, GeneAmp®, PCR Instrument System 9700, USA).

Os produtos da amplificação foram aplicados em gel de agarose a 1.0 %, corados com GelRed e revelados em eletroforese horizontal com tampão TBE 0,5 X e Ladder de 1 kb (Invitrogen), conduzida em 90 V por 1 h. Os amplicons foram visualizados em transiluminador sob luz ultravioleta, fotodocumentados e verificado a presença de uma banda de aproximadamente 300 pb que corresponde a região de interesse do gene *psy2*.

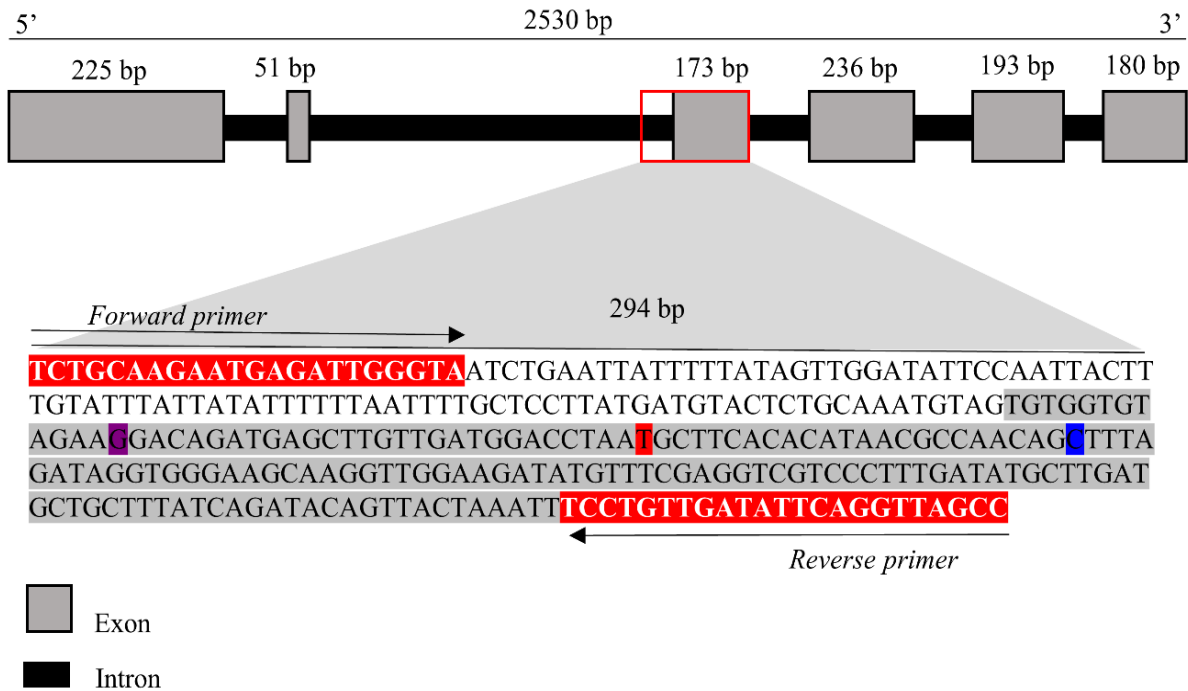


Figura 2.1. Representação esquemática do gene fitoeno sintase 2 (*psy2*) de mandioca (*Manihot esculenta*), mostrando as regiões de ligação dos primers *forward* e *reverse*. O éxon está destacado em cinza, as regiões de ligação dos primers estão destacadas em vermelho e as letras pretas sem preenchimento pertencem a uma região de íntron.

2.2.4 Purificação e sequenciamento do DNA

Os produtos da PCR, em que as amostras continham aproximadamente 300 pb foram purificados usando o kit de purificação PureLink Quick Gel Extraction (Invitrogen, Brasil), seguindo recomendações do fabricante, e em seguida as amostras com concentração entre 30 a 70 ng μL^{-1} de DNA, confirmados por eletroforese, foram sequenciadas pela Helixxa, Brasil, utilizando sequenciamento automático pela plataforma Applied Biosystems® 3500 Genetic Analyzer.

2.2.5 Análise de bioinformática

Os eletroferogramas das sequências “*forward*” e “*reverse*” gerados foram inspecionados visualmente por meio do programa BioEdit versão 7. 2. Após obtenção das sequências, foi feito BLAST da sequência de nucleotídeo para comparações com outras sequências disponíveis no GenBank.

Sequências consensos, contigs, para cada acesso foram obtidas com base nas fitas F e R utilizando o programa CAP3 (Huang e Madan 1999). Os *contigs* foram alinhados com a sequência referência *Manihot esculenta* cultivar CM3306-4 gene *psy2*, alelo PSY2-W-1

(GenBank: GU111715). A Coding Sequence (CDS) foi utilizada para obtenção da sequência de aminoácidos deduzida utilizando o software MEGA 6.0 (Tamura et al. 2013) e construídas as redes de haplótipos das sequências de nucleotídeos e de aminoácidos da região codificante de 173pb (éxon 3) usando o software Haploview (Salzburger et al. 2011).

2.2.6 Modelagem molecular

Para verificar o efeito dos polimorfismos não sinônimos (SNP1 e SNP3) identificados na proteína PSY2, foi realizada a modelagem de quatro sequências proteicas, incluindo a sequência referência ACY42665.1 que é semelhante as sequências sem polimorfismo do presente estudo. As sequências homólogas que foram usadas para identidade de resíduos conservados foram obtidas usando BLASTP (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Assim foi feito o BLASTP de acesso contendo apenas o SNP1, em que ocorre troca do aminoácido Arg por Thr; BLASTP de acesso apenas com o SNP3, que ocorre a troca de uma Ala por um Asp; e BLASTP de acesso que apresentaram os SNP1 e SNP3 na mesma sequência; também foi realizado o BLASTP de sequência sem polimorfismo, sendo igual a sequência de aminoácido da sequência referência.

Os modelos para os acessos com SNP1 e/ou SNP3 foram selecionados usando o programa HHpred (Soding et al., 2005). Cinco modelos foram utilizados para as quatro proteínas: as estruturas dos arquivos PDB 4hd1, 3acx, 3we9, 3vj8 e 3wca. A estrutura protéica de três acessos com polimorfismo e uma referência foi obtida usando Modeller (versão 9.13) (Sali e Blundell 1993). Em seguida, os acessos modelos produzidos pelo Modeller tiveram suas estruturas de energia minimizadas usando o GROMACS 4.6.5. (Berendsen et al., 1995) com o campo explícito solvente (modelo spc) e AMBER99SB-ILDN. Brevemente, para começar a minimizar a energia e melhorar os contatos da van der Waals, uma dinâmica de posição restrita foi realizada usando um algoritmo de descida mais íngreme e definindo um modelo de água SPC flexível (define = -DFLEXIBLE; total de 100 ps). A próxima posição de minimização de energia restringiu a macromolécula e permitiu que as moléculas de água entrassem (definir = -DPOSRES; restrições = todas as ligações; integrador = md; total de 20 ps). A última dinâmica de minimização de energia não restringiu as posições (integrador = md; total de 50 ps). O programa VMD (VMD para LINUXAMD64, versão 1.9.1) (Humphrey et al., 1996) foi usado para criar as imagens do modelo molecular e para calcular o RMSD (Root mean square deviation) dos acessos em relação à proteína de referência acy42665.1. Para isso, os quatro acessos foram alinhados em relação à proteína de referência, usando

apenas os resíduos 133-416 para evitar a alta variação encontrada nas sequências próximas dos terminais C e N. A RMSF (Root mean square fluctuation) por resíduo em relação ao tempo foi calculada com GROMACS.

2.3 Resultados

O alinhamento das sequências de DNA de 78 acessos de *Manihot esculenta* com a sequência referência GU111715 revelou a existência de oito haplótipos e foi observado que 56 acessos apresentaram de uma a três substituições de nucleotídeos em homozigose ou heterozigose que correspondem aos SNP1, SNP2 e SNP3 localizados nas posições 1434, 1462 e 1485 respectivamente; e 22 acessos não apresentaram substituição de nucleotídeos, sendo semelhantes a GU111715 (Tabela 2.2), apresentando 100% de identidade.

De acordo com a rede de haplótipos baseada na sequência de nucleotídeos (Fig. 2.2a) e Tabela 2.2, os acessos foram organizados em oito grupos, que correspondem aos oito haplótipos (Fig. 2.2b). O grupo PSY2-A1 apresentou 18 acessos com raiz branca (W), todos compartilhando a substituição da timina (T) por uma citosina (C) na posição 1462 (SNP2, c.1462T >C). O haplótipo PSY2-A2 com 3 acessos e fenótipo apenas de raiz amarela apresentou os SNP1 e SNP2 que corresponde as trocas de base guanina (G) por citosina (C) e T por C respectivamente (SNP1, c.1434G>C; SNP2, c.1462T>C) e está associado ao haplótipo PSY-A3.

Tabela 2.2. Identificação dos SNPs na região do éxon 3 do gene da fitoeno sintase 2 (*psy2*) e em 78 acessos de mandioca da Amazônia e seus respectivos haplótipos.

Haplótipos	SNP1 c.1434 G>C	SNP2 c. 1462 T>C	SNP3 c.1485 C>A	SNP	Nº de acessos	Cor da raiz
GU111715	G	T	C	-	Referência	Branca
PSY2-A1	-	C	-	1	18	Branca
PSY2-A2	C	C	-	2	3	Amarela
PSY2-A3	C	-	-	1	10	Amarela e creme
PSY2-A4	-	-	-	0	22	Branca
PSY2-A5	-	-	A	1	1	Creme
PSY2-A6	-	C	A	2	8	Amarela e creme
PSY2-A7	C	-	A	2	3	Amarela
PSY2-A8	C	C	A	3	13	Amarela

O PSY-A3 com 10 acessos foi formado por nove acessos com fenótipo de raiz amarela e um de raiz creme, todos apresentando o SNP1 (c.1434G>C). O grupo de haplótipos com 22 acessos com apenas raiz branca, não apresentou mudança de nucleotídeos em relação a

sequência de referência. O haplótipo PSY2-A5 apresentou tonalidade de raiz creme e apresentou apenas o SNP3 que corresponde a mudança de base citosina (C) por uma adenina (A) (c.1485C>A). Esse grupo com apenas um acesso está ligado ao grupo PSY2-A7 (Fig. 2.2a).

Com oito acessos, o haplótipo PSY2-A6 foi formado por seis acessos de raiz amarela e dois acessos de raiz creme, em que todos apresentaram os SNP2 e SNP3, que correspondem as mudanças de base c.1462T>C e c.1485C>A, respectivamente. PSY2-A7 foi formado por três acessos de raiz amarela e apresentou os SNP1 e SNP3 que correspondem a mudança de base c.1434G>C e c.1485C>A respectivamente.

Formado apenas por acessos com raízes de tonalidade amarela, o haplótipo PSY2-A8 apresenta os SNPs 1, 2 e 3 (c.1434G>C, c.1462T>C e c.1485C>A) (Tabela 2.2).

Portanto, foi observado que seis grupos haplótipos foram formados por acessos de tonalidade creme e/ou amarela, e apresentaram os SNP1 e/ou SNP3 que foram associados com a cor da raiz de mandioca.

Para verificar se os polimorfismos no DNA produzem mudanças na sequência de aminoácidos da proteína PSY2 ou apenas ‘silent changes’, as sequências de DNA contendo os SNPs identificados na população foram deduzidas e analisadas. A rede de haplótipos baseada na sequência de aminoácidos é diferente da rede de haplótipos baseada na sequência de nucleotídeo, principalmente porque a mutação do SNP2 (c.1462T>C) é silenciosa e não promove mudança nas sequências de aminoácidos.

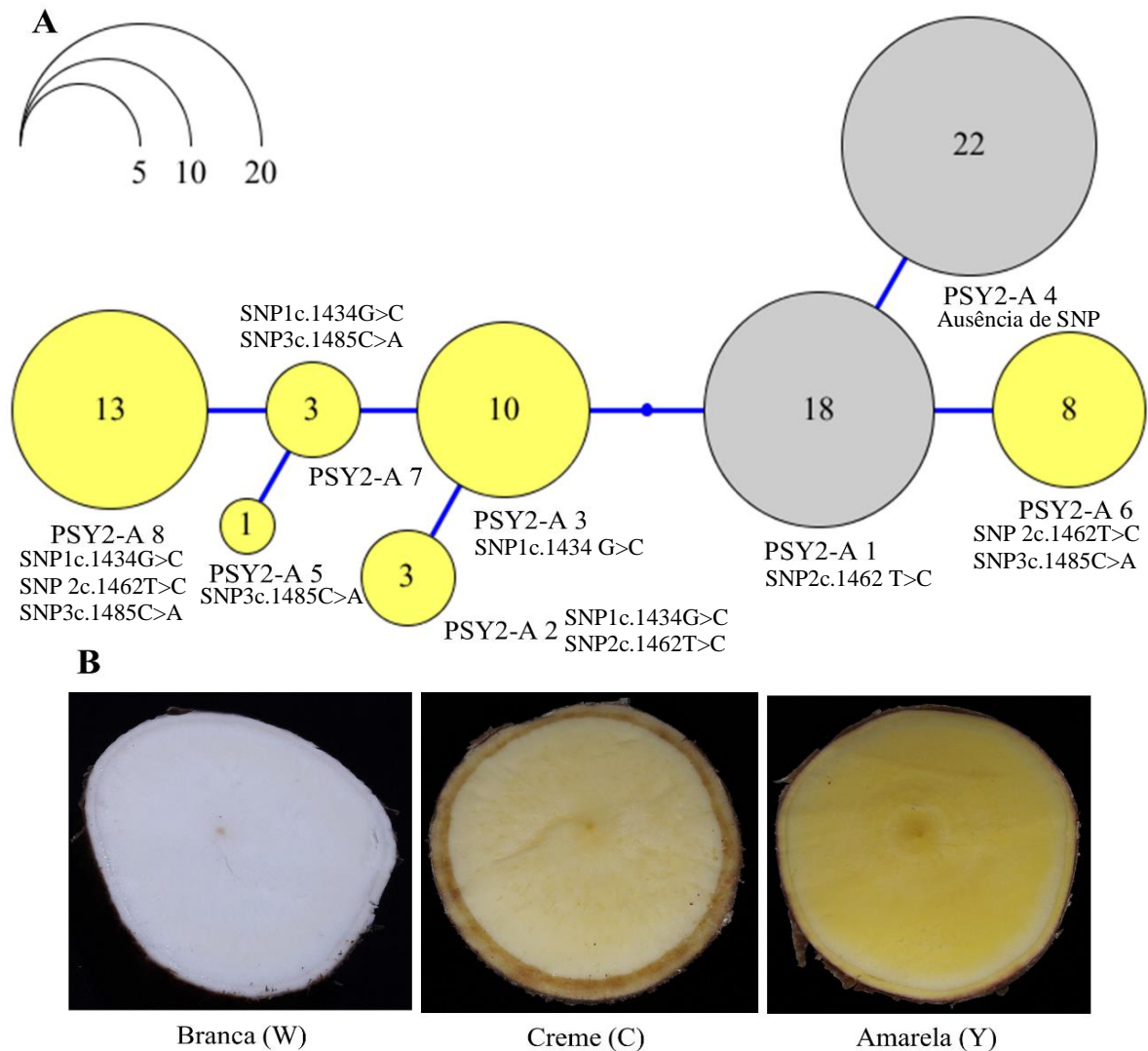


Figura 2.2. Rede de haplótipos de acessos de mandioca (*Manihot esculenta* Crantz) com tonalidade branca, creme e amarela. **a** Rede de haplótipo construída com base nas sequências de 173 pb do éxon 3 do gene da fitoeno sintase 2 (*psy2*) de 78 acessos de mandioca obtida pelo programa Haploview. A cor cinza dos círculos representa acessos com raiz de tonalidade branca, cor amarela dos círculos representa acessos com raiz de tonalidade amarela ou creme. O tamanho do círculo é equivalente ao tamanho da amostra. **b** Representação das raízes de mandioca branca, creme e amarela.

A rede de haplótipos baseada na sequência de aminoácidos foi composta por quatro grupos (Fig. 2.3). O grupo de haplótipos com 40 acessos foi composto apenas por acessos com raiz branca, nenhum acesso desse grupo apresentou mudança de aminoácido, visto que estes possuem o genótipo selvagem ou apresentam apenas o SNP2, que é uma mutação silenciosa, sendo semelhante a sequência de aminoácido referência (ACY42665.1). O grupo com 13 acessos foi formado por 12 plantas com fenótipo de raiz amarela e uma de raiz com tonalidade creme, e nesse grupo, os genótipos apresentaram o SNP1 que equivale a mudança de aminoácido arginina (Arg) por treonina (Thr) na posição 174 da proteína (p.Arg174Thr)

que equivalem a sequência ACY42669.1 disponível no NCBI. O grupo de haplótipo com nove acessos foi formado por seis plantas de raiz amarela e três de raiz creme, em que todos apresentaram o SNP3 que corresponde a mudança do aminoácido alanina (Ala) por ácido aspártico (Asp) na posição 191 (p.Ala191Asp) e são equivalentes a sequência ACY42670.1. O grupo com 16 acessos foi formado por 16 plantas de raiz com totalidade amarela e foram os únicos que apresentaram os SNP1 e SNP3 na mesma sequência de DNA, que correspondem as mudanças dos aminoácidos p.Arg174Thr e p.Ala191Asp respectivamente (Fig. 2.3). Os acessos desse grupo não foram semelhantes a sequências disponíveis no NCBI.

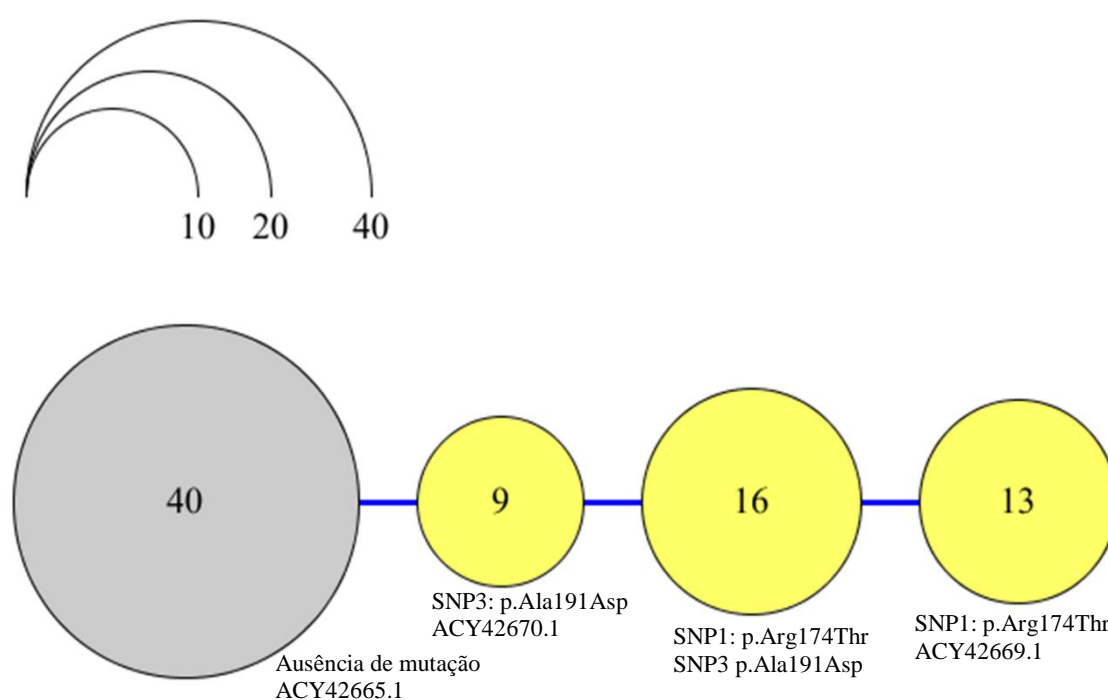


Figura 2.3. Rede de haplótipos construída com base nas sequências de aminoácidos do exon 3 do gene que codificadora a proteína fitoeno sintase 2 (PSY2) com base em 78 acessos de mandioca da região Amazônica. A rede de haplótipo foi obtida pelo programa Haploview. O círculo cinza apresenta haplótipos com raiz de tonalidade branca, os círculos amarelos apresentam haplótipos com raiz de tonalidade amarela e creme. O tamanho do círculo é equivalente ao tamanho da amostra.

Os SNP1 e SNP3 se destacaram por estar presentes em acessos de raízes com tonalidade amarela e creme, e por representarem polimorfismo em que a substituição de base equivale a uma troca de aminoácido (Fig. 2.4a). O SNP2 foi encontrado em acessos com raízes de cor branca, não sendo associado, portanto ao fenótipo amarelo ou creme. Assim, os acessos que apresentaram apenas o SNP2 são semelhantes ao selvagem, o que justifica a inclusão destes no grupo de haplótipo de 40 acessos (Fig. 2.3).

O alinhamento das sequências de aminoácidos mostrou que as plantas produtoras de raiz de cor branca não apresentaram mudança de aminoácido, sendo semelhante a sequência referência GU111715 com síntese dos aminoácidos Arginina (Arg) na posição 174 e Alanina (Ala) na posição 191. Entretanto, nos acessos de raiz amarela e creme, ocorreram trocas de um ou dois aminoácidos, pois o aminoácido Arg foi substituído por Thr e Ala substituído por Asp. Esses resultados mostram a relação positiva entre os SNPs e a cor da raiz de mandioca (Fig. 2.4a).

Todos os acessos de raiz branca estão em homozigose na posição dos SNP1 e SNP3 e não apresentaram troca de nucleotídeo e aminoácido, sendo semelhantes a referência GU111715. Sete acessos de raiz amarela apresentaram homozigose para o SNP1 (Fig. 2.4a) e dos 38 acessos com tonalidade amarela ou creme apenas um acesso de raiz amarela apresentou homozigose no SNP3, que foi o acesso CPATU 018, pertencente ao grupo de haplótipo PSY2-A6 (Tabela 2.1 e Fig. 2.2a). Quatro acessos apresentaram homozigose para o SNP2, porém esse SNP não substituiu aminoácido (Fig. 2.4a-b). Portanto, foi verificado que entre os acessos sequenciados, o fenótipo cor da raiz branca se manifestou apenas em homozigose e o fenótipo cor de raiz amarela ou creme não necessitou estar em homozigose para se manifestar (Fig. 2.4b).

Os haplótipos PSY2-A3, PSY2-A6 e PSY2-A8 estão mais amplamente distribuídos nos estados do Pará, Amapá, Rondônia e Amazonas (Fig. 2.5), e foram detectados em acessos que apresentaram cor de polpa da raiz creme e amarela. Os haplótipos PSY2-A1 e PSY2-A4 estão distribuídos nos Estados do Pará e Amapá, porém não apresentaram mudança de aminoácidos, sendo acessos predominantemente com raiz branca. O haplótipo PSY2-A2 predominou na região do Baixo Amazonas do Estado do Pará e foi detectado apenas em acessos com raiz amarela. O haplótipo PSY2-A7 está presente apenas em acessos de raiz amarela e foi distribuído no Estado do Pará nas regiões do Baixo Amazonas, Arquipélago do Marajó e Região Tocantins. O haplótipo PSY2-A5 foi encontrado no Estado do Amapá, sendo representado por apenas um acesso com cor de raiz creme, porém esse haplótipo apresentou polimorfismo que corresponde a uma troca de aminoácido que está associado a cor amarela da raiz de mandioca (Fig. 2.5).

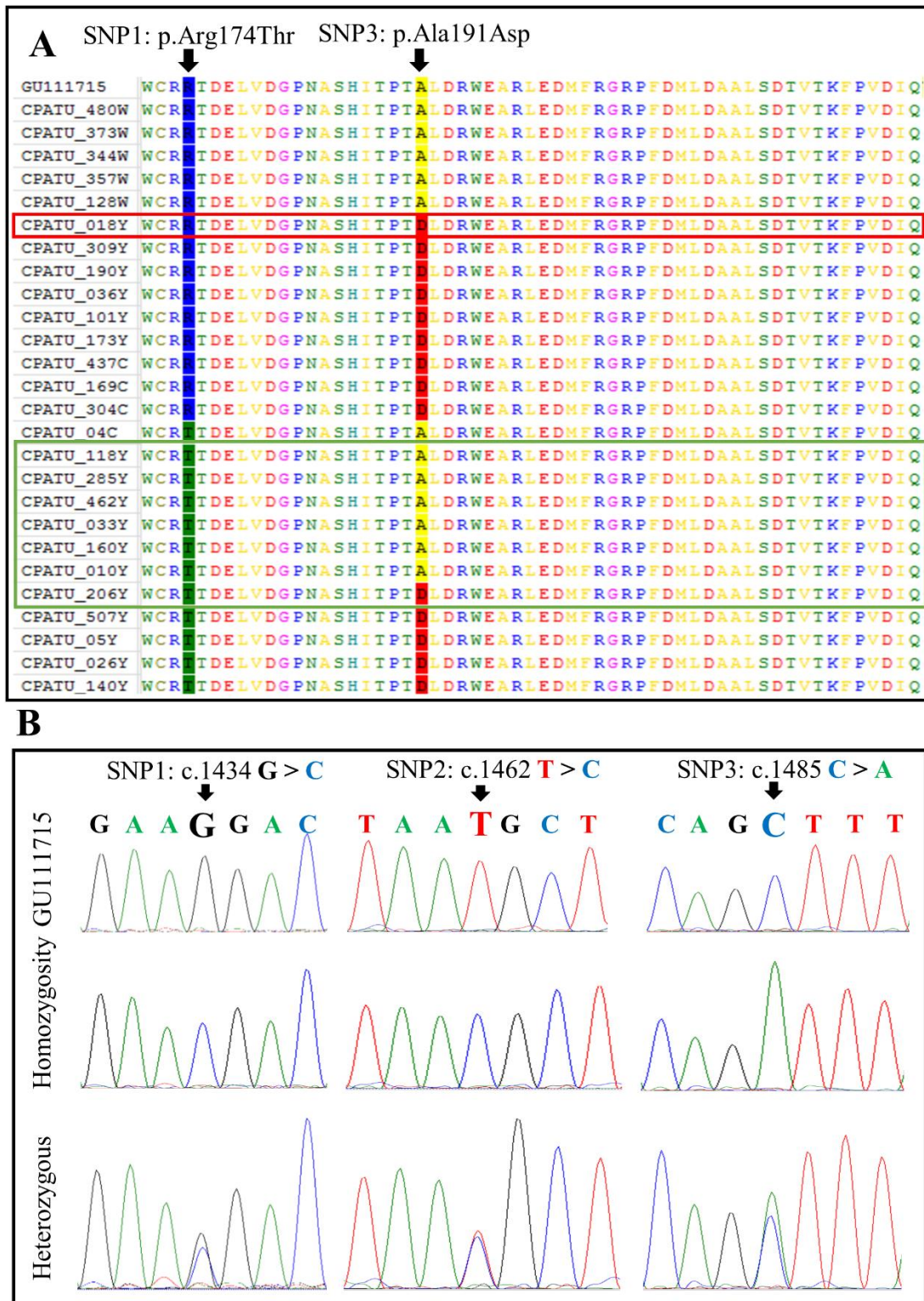


Figura 2.4. Variantes da proteína e do gene da fitoeno sintase 2 em acessos de mandioca (*Manihot esculenta*) com raízes de diferentes tonalidades. **a** Alinhamento das sequências de aminoácidos de acessos com tonalidade de raiz branca (W), creme (C) e amarela (Y), retângulo verde acessos com homozigoze para o SNP1, retângulo de cor vermelha representa acesso com homozigoze para o SNP3. **b** Picos do cromatograma de SNPs em homozigoze e heterozigoze da região funcional do gene fitoeno sintase 2 (*psy2*).

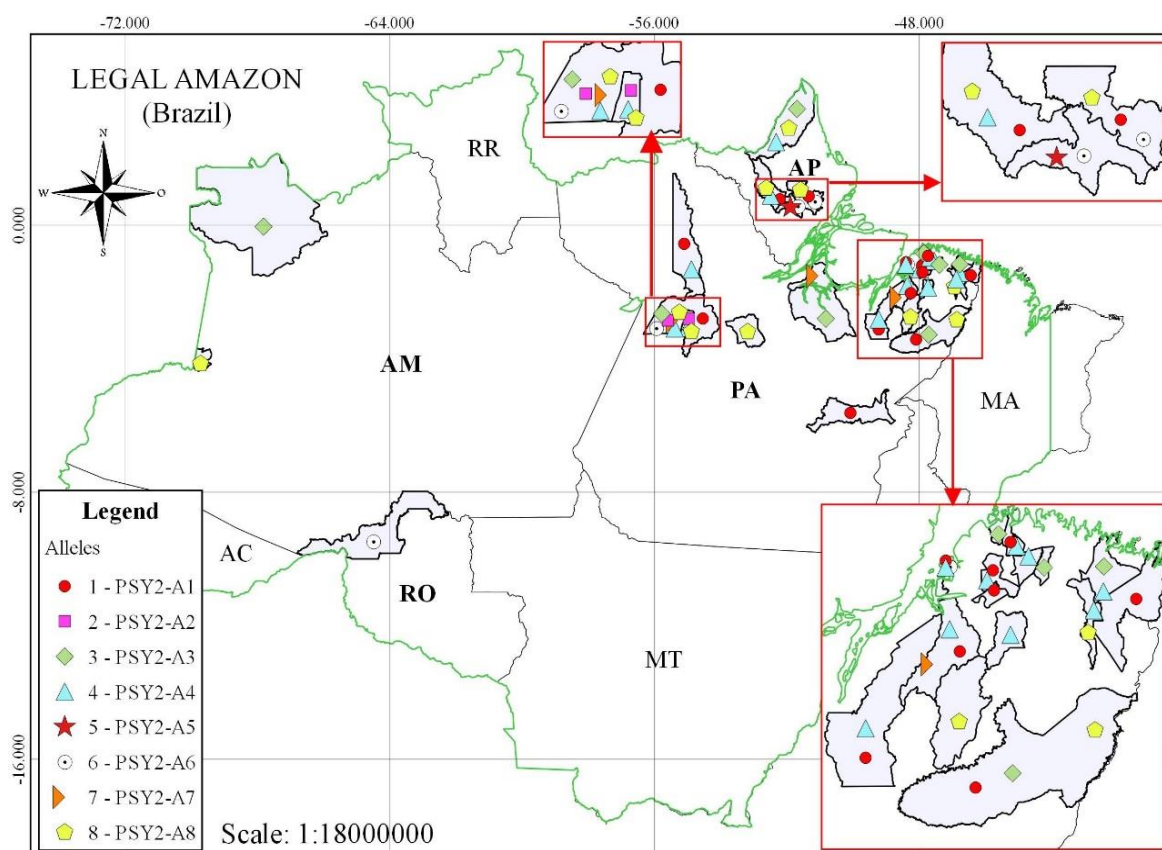


Figura 2.5. Distribuição geográfica dos oito haplótipos do gene fitoeno sintase 2 (*psy2*) de mandioca (*Manihot esculenta* Crantz) identificados após o sequenciamento de fragmento de 300 pb dos 78 acessos originários da Amazônia brasileira.

Para ajudar a verificar os efeitos das mutações nos resíduos 174 e 191 (SNP1 e SNP3) na estrutura da proteína e no sítio ativo, os resíduos mais conservados dentro da família do esqualeno e da fitoeno sintase foram identificados. Usando as suas posições FASTA, os resíduos selecionados foram F150, Y169, R173, D176, D180, Y254, V258, A259, V262, G263, G291, Q295, N298, I299, R301, D302, D306 e R311 (Pandit et al. 2000). O modelo molecular preditivo de acy42665.1 é mostrado na Fig. 2.6a. Quando os resíduos mais conservados são destacados com o método de desenho VMD *surf*, um núcleo de proteína emerge com uma cavidade clara que se assemelha a um sítio ativo. E os resíduos encontrados dentro da cavidade são compatíveis com sua frequência usual em enzimas (Pravda et al. 2014). Por exemplo, os resíduos de Gly são conhecidos por fornecer flexibilidade para sítios ativos (Yan e Sun 1997) e são frequentemente encontrados no meio de canais enzimáticos (Pravda et al. 2014). E dois resíduos de Gly muito bem conservados, Gly-263 e Gly-291 (Fig. 2.6b), são encontrados dentro da cavidade do sítio ativo e apresentaram RMSD abaixo da média.

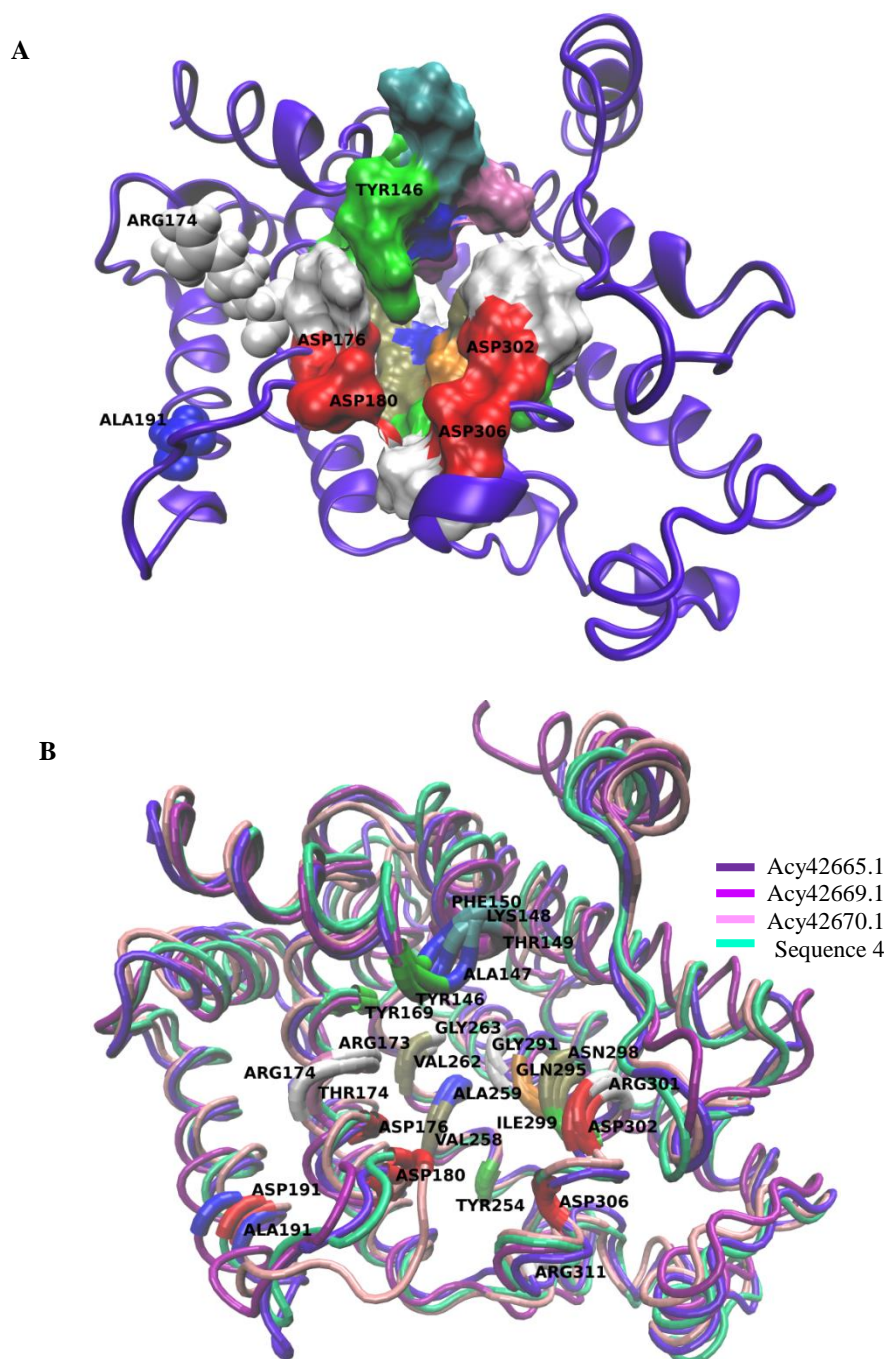


Figura 2.6. Modelos preditivos de fitoeno sintase. Para maior clareza, apenas os resíduos 133-416 são mostrados. Os resíduos mais conservados entre a família de esqualeno e fitoeno sintase, bem como a região importante 146-149, e os mutantes 174 como 191 são coloridos pelo nome do resíduo. Os arredores são coloridos por cadeias. **a** Uma representação em 3D do acy42665.1. Os resíduos conservados formam um núcleo com uma cavidade, onde o sítio ativo presumido está localizado. Os resíduos 174 e 191 foram destacados com o raio Van der Waals e parecem estar separados da superfície da cavidade. **b** Um alinhamento 3D dos quatro mutantes, usando acy42665.1 como referência. Os dois motivos ricos em ASP₁₇₆DELVD₁₈₀ e ₃₀₂DVGED₃₀₆ se enfrentam na região externa do canal. Os resíduos 146-150 correspondem à "flap" flexível de SQS- Esqualeno sintase (resíduos 50-54), que se pensa proteger da água os intermediários de reação (Umeno et al 2002, Pandit et al., 2000).

Visualmente, os resíduos mutantes 174 e 191 não pareciam causar grandes mudanças na cavidade. Quando o suposto sítio ativo de quaisquer dos 4 acessos é observado de dentro de sua cavidade, com o método *surf* de desenho VMD, o resíduo 174 é difícil de ver e parece estar fora do alcance (dados não mostrados). Além disso, o efeito dessas mutações na estrutura 3D das proteínas foi avaliado tomando o RMSD da cadeia principal, entre as seqüências. Acy42665.1 foi utilizado como referência, e cada uma das outras 3 seqüências (acy42669.1, acy42670.1 e seqüência 4) foram emparelhadas com a referência, formando 3 pares (Fig. 2.7). As proteínas foram alinhadas usando apenas os resíduos 133-415 para evitar as altas variações próximas ao N-terminal e C-terminal. O RMSD médio total foi de 1,88 Å. Não surpreendentemente, os valores de RMSD mais elevados foram geralmente encontrados dentro ou ao lado de loops, que são intrinsecamente mais flexíveis (Liu et al. 2002) para poder desempenhar suas funções (Kempner 1993).

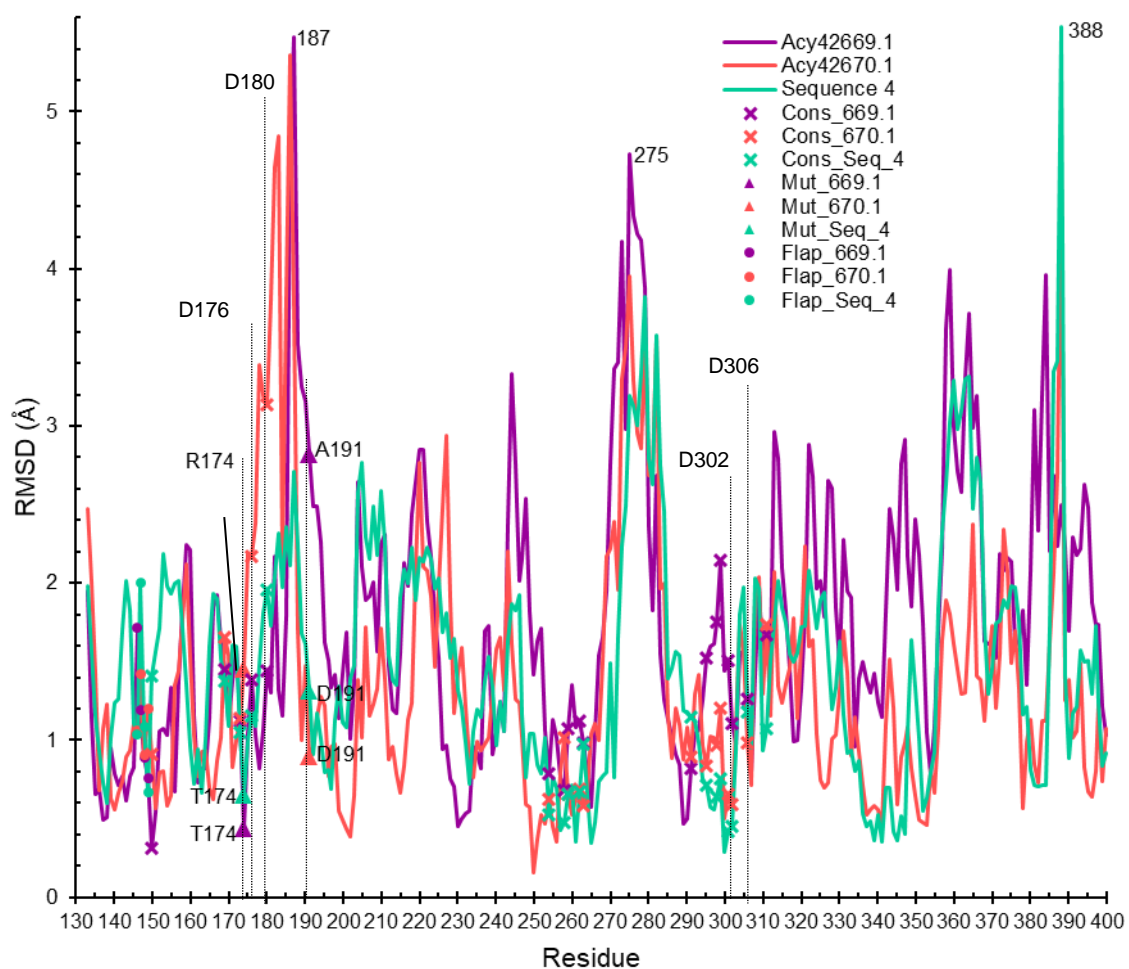


Figura 2.7. O RMSD por resíduo entre os mutantes dos acessos. Acy42665.1 é a referência. Os resíduos 174 e 191 são os mutantes, e na seqüência de referência são R174 e A191. Os valores mais altos tendem a estar dentro das regiões com mais características de alças. Os resíduos Asp D176, D180, D302 e D306 são altamente conservados. A mutação do resíduo 302 inativa a enzima.

O RMSD médio de acy42669.1, 42670.1 e sequência 4, em relação à referência foram 2.2, 1.8 e 1.7, respectivamente. RMSDs em torno de 2 Å indicam diferença significativa entre estruturas terciárias (Carugo 2007). Esse RMSD inferior apresentado pela sequência 4 sugere que as mutações duplas podem resultar em diminuição da RMSD em comparação com uma única mutação no resíduo 174 (acy42669.1) ou 191 (acy42670.1). Do mesmo modo, pensa-se que as mutações duplas restauram a aptidão para HIV-1 em comparação com mutantes de base única (De Conto et al., 2015). O único resíduo conservado com RMSD superior a 2 Å em relação à referência foi I299 de acy42669.1, e os aspartatos 176 e 180 de acy42670.1 (RMSD = 2,14, 2.17 e 3.1, respectivamente).

2.4 Discussão

O crescimento acelerado da população mundial tem desafiado a produção agrícola e este fato tem exigido maiores esforços para otimizar o uso de recursos necessários para o aumento da produtividade com a redução de efeitos ambientais (Moose e Mumm 2008). A mandioca possui uma grande diversidade genética em caracteres agrícolas importantes para o fornecimento de alimento para este rápido aumento da população humana (Ceballos et al. 2015; Nassar et al. 2007; Nassar 2003). Dentro deste contexto, é importante salientar que um único polimorfismo nucleotídico no gene que codifica a proteína *PSY2* leva a aumento acentuado da formação e acumulação de carotenoide nas raízes de mandioca, o qual é precursor da vitamina A (Welsch et al. 2010), e a confirmação desse SNP como marcador molecular em diferentes background genéticos, como apresentado no presente estudo, pode representar a identificação de ferramenta para acelerar os programas de melhoramento genético da mandioca. Além disso, tecnologias de edição do DNA como o CRISPR/CAS9 podem levar ao melhoramento direcionado de genótipos já adaptados a região de interesse, sem necessidade de envolver cruzamentos recorrentes que demandam tempo e custos.

No presente estudo, seis grupos haplótipos (Fig. 2.2a) são formados por acessos que apresentam coloração amarela ou creme e apresentam marcador do tipo SNP, em que ocorre a mudança de nucleotídeos que podem ocasionar a mudança na sequência de aminoácidos na proteína *PSY 2* (Fig. 2.4a). Essa variação de aminoácidos na sequência proteica e a formação de quatro grupos haplótipos apresentado na Fig. 2.3, mostra que o fenótipo cor da raiz pode ser associado a presença de dois polimorfismos.

No presente estudo, o gene *psy2* apresentou três SNPs (SNP1, SNP2 e SNP3) em uma região codificadora, sendo os SNP1 e SNP3 associados com o fenótipo cor da raiz de

mandioca creme ou amarela, pois entre os 78 acessos estudados, 38 incluindo raízes amarela e creme apresentaram os SNP1 e/ou SNP3, com substituição de G (purina) por C (pirimidina) que equivale a troca do aminoácido p.Arg 174Thr e de C (pirimidina) por A (purina) que equivale a troca do aminoácido p.Ala191Asp, respectivamente (Fig. 2.2a e 2.4a). Essas mutações são do tipo transversão, não sinônimas e promoveram a substituição de aminoácido. Nesse caso, a presença do polimorfismo pode ter levado a uma mudança estrutural da proteína, e conseqüentemente a uma possível alteração da sua função (Stitzel et al. 2004), o que reforça que mesmo uma pequena diferença na sequência de aminoácidos da proteína PSY2 pode ter grandes efeitos em termos da estrutura e dobramento da proteínas, assim como em suas propriedades enzimáticas (Arango et al. 2010; Paine et al. 2005), com isso os acessos que apresentam mudanças não sinônimas (Tabela 2.2 e Fig. 2.3), são recomendados para serem usados em trabalho de melhoramento genético da mandioca, pois apresentam SNP associado ao fenótipo cor da raiz.

Os resultados encontrados no presente estudo confirmam os encontrados por Welsch et al. (2010), onde foi demonstrado que o SNP3 proporciona a substituição p.Ala191Asp (Fig. 2.4a), sendo importante para o fenótipo cor amarela da raiz mesmo estando em heterozigose, o que reforça a comprovação da dominância desse alelo. A dominância do SNP3 pode produzir o dobro de fitoeno e aumentar a atividade catalítica da enzima PSY2 em raiz com tonalidade amarela e com isso aumentar a concentração de carotenóides pró-vitamina A na raiz de mandioca (Welsch et al. 2010).

É importante destacar que o SNP3 foi encontrado de forma isolada em apenas um acesso (PSY2-A5), e de forma conjunta com os SNP1 e SNP2 nos grupos haplótipos PSY2-A5, PSY2-A6, PSY2-A7 E PSY2-A8 (Fig. 2.2a), todos com raiz de cor creme ou amarela. Esse resultado reforça a importância do SNP3 para o fenótipo cor da raiz de mandioca. Liu et al. (2015) estudando a diversidade e distribuição do gene puroindoleno-D1 em trigo, encontraram resultados semelhantes, em que a troca de nucleotídeo pode não corresponder a troca de aminoácido. Isso justifica o fato da rede de haplótipos baseada nos nucleotídeos ser diferente da baseada nos aminoácidos (Fig. 2.2a e 2.3). A identificação do tipo de mutação no gene é importante porque mutações não sinônimas, iguais às encontradas no presente estudo no SNP1 e SNP3, são biologicamente mais relevantes por resultar em troca de aminoácidos e por isso interferem na estabilidade, estrutura, solubilidade e função das proteínas, podendo resultar em mudanças no fenótipo (Kim et al. 2003; Tang et al. 2006). Fu et al. (2010), estudando a diversidade de nucleotídeos e evolução do gene *psy1* em milho, descreveram substituições não sinônimas sendo importantes para a separação do milho amarelo e branco.

Apesar de ainda há muito o que saber sobre os fitoenos, várias intuições podem ser derivadas sobre os fitoenos mutantes, comparando-o com proteínas relacionadas. As sequências $_{176}\text{DELVD}_{180}$ e $_{302}\text{DVGED}_{306}$ são conservadas entre as isoprenóides sintases (Shumskaya et al., 2012). Imagina-se que as regiões correspondentes em Esqualeno sintase (SQS), $_{80}\text{DTLED}_{84}$ e $_{219}\text{DYLED}_{223}$ ligam difosfatos de dois substratos (Pandit et al. 2000). A mutagênese de Asp-285 (D302) inativa tanto SQS quanto Zm-PSY1 (Shumskaya et al. 2012).

Os muito bem conservados resíduos hidrofóbicos Val-258 e Ala-259 fazem parte da base da fenda. Ambos apresentaram RMSD abaixo da média em todos os quatro acessos. A hélice forma uma torção (*kink*) neste segmento, como observado em SQS (nos resíduos correspondentes Val-175 e Ala-176 (Pandit et al., 2000). Alfas-hélices formam ligações de hidrogênio entre os grupos amino e ácido carboxílico da cadeia principal dentro da hélice (Pauling et al. 1951). Portanto, pode-se esperar que essa mudança na direção da hélice exponha esses grupos à formação de ligações de hidrogênio com outros ligantes, e já foi proposto que essa torção tem um papel funcional em SQS (Pandit et al. 2000).

A chance de destruir a atividade da enzima por uma mutação é muito mais curta na superfície do que no núcleo (Burg e Eijnsink 2002), e os resíduos mutantes 174 e 191 estão ambos localizados na superfície. Assim, as mudanças provocadas pelas propriedades do novo resíduo tendem a ser mais suaves. A mutação de Ala-191 (como em acy42665.1) em Asp-191 (como na sequência 4) deve ter pouco efeito na flexibilidade, já que Ala e Asp têm flexibilidade semelhante. Em contraste, uma mudança de Arg-174 (como em acy42665.1) para Thr-174 (como na sequência 4) pode resultar em efeitos de flexibilidade, uma vez que Thr é mais flexível do que Arg (Huang e Nau 2003). Em harmonia com isso, os valores médios de RMSF (Å) dos resíduos mutantes tendem a seguir a escala de flexibilidade: Asp-191 = Ala-191 > Thr-174 > Arg-174 (RMSF médio = 0,06, 0,066, 0,059 e 0,048, Respectivamente).

O RMSD médio de resíduos conservados em comparação com a referência foi 1,30, 1,31 e 1,01 para acy42669.1, acy42670.1 e sequência 4, respectivamente. O significado desse tipo de diferenças é difícil de determinar e é discutido (Carugo 2007). No entanto, corroboram a diferença de cor encontrada nestes 3 mutantes: os únicos mutantes que diferiram estatisticamente em cores foram acy42670.1 e sequência 4 (Teste Exato de Fisher, dois lados, $P < 0,05$).

O RMSD dos resíduos conservados indicou algum grau de mudanças conformacionais, principalmente nos altamente conservados D176 e D180 de acy42670.1, ambos próximos da mutação $_{\text{ALA}}191_{\text{ASP}}$. Isso sugere que essas mutações afetam o sítio ativo à

distância, resultando em rearranjos leves nas posições espaciais relativas dos resíduos ativos. As alterações combinadas provavelmente modificarão a taxa de reação catalisada, talvez em vários passos, o que poderia envolver não só o próprio sítio catalítico, mas também flexibilidade da enzima e a filtragem de moléculas aceitas pelo canal (Kingsley e Lill 2015), especialmente por guardiões (gate-keepers) (Pravda et al. 2014). Ou talvez até a ligação de inibidores, talvez em Phe150 ou Tyr169 (Fig 6a), conhecida por formar duas paredes da cavidade de ligação de inibidor nos seus equivalentes em SQS, os resíduos Phe54 e Tyr73, respectivamente (Pandit et al. 2000). Todas essas possibilidades precisam de mais investigações.

Portanto a identificação de novos haplótipos pode ser importante para as formas de utilização da raiz de mandioca, pois acessos de mandioca de raiz de cor branca, podem ser indicados para uso na indústria de amido, farmacêutica, cosméticos e produtos biodegradáveis, enquanto os genótipos de raízes com tonalidade amarela ou creme podem ser usados diretamente na alimentação, através do preparo de comidas típicas, com o objetivo de suprir deficiência de vitamina A. Assim novos haplótipos podem fornecer recursos genéticos para modificar a cor da raiz de mandioca por meio de alterações em regiões que influenciam a expressão do gene *psy2*.

A identificação e distribuição geográfica dos haplótipos PSY2-A5, PSY2-A6, PSY2-A7 e PSY2-A8 em que todos os acessos apresentaram mudança de base c. 1485C>A no SNP3 (Fig. 2.5) e mudança de aminoácido é importante, pois as origens dos recursos genéticos da mandioca para uso potencial estão em diferentes Estados da região amazônica como nos Estados do Pará, Amapá, Rondônia e Amazonas, além de estarem localizados em diferentes regiões do Estado do Pará, onde está localizado o BAG, sendo o Estado do Pará um importante sítio de coleta para preservação e estudo de acessos originários da região Amazônica brasileira.

2.5 Conclusão

As variações alélicas do gene *psy2* que apresentaram os SNP1 e SNP3, com substituição dos aminoácidos p.Arg174Thr e p.Ala191Asp respectivamente, foram predominantes em acessos de mandioca de raiz com tonalidade amarela e creme, mostrando que a presença do polimorfismo na região do éxon 3 está associada com tonalidade da raiz de mandioca e apresentam uma ampla distribuição em acessos coletados na região amazônica. Os dados de modelagem da proteína indicam que as substituições de aminoácidos afetam o

sítio ativo conforme a distância, resultando em rearranjos leves nas posições espaciais relativas dos resíduos ativos.

O BAG forneceu recursos genéticos, com uso potencial para trabalhos de melhoramento genético, visando aumentar a concentração de carotenóides pró-vitamina A em raiz de mandioca originárias da Amazônica brasileira. E o acesso CPATU 018, por está em homozigose para o SNP3, é o mais indicado para iniciar os trabalhos de melhoramento do BAG da Amazônia Oriental.

REFERÊNCIAS

- Arango J, Jourdan M, Geoffriau E, Beyer P, Welsch R (2014) Carotene Hydroxylase Activity Determines the Levels of Both-Carotene and Total Carotenoids in Orange Carrots. *The Plant Cell* 26:2223–2233. <https://doi.org/10.1105/tpc.113.122127>
- Arango J, Wust F, Beyer P, Welsch R (2010) Characterization of phytoene synthases from cassava and their involvement in abiotic stress-mediated responses. *Planta* 232:1251–1262. <https://doi.org/10.1007/s00425-010-1250-6>
- Bang H, Kim S, Leskovar D, King S (2007) Development of a codominant CAPS marker for allelic selection between canary yellow and red watermelon based on SNP in lycopene β -cyclase (LCYB) gene. *Molecular Breeding* 20:63–72. <https://doi.org/10.1007/s11032-006-9076-4>
- Berendsen HJC, Van der Spoel D, van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* 91:43–56. [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E)
- Brookes AJ (1999) The essence of SNPs. *Gene* 234:177–86. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10395891>
- Burg VB and Eijsink VG (2002) Selection of Mutations for Increased Protein Stability. *Current Opinion in Biotechnology* 13:333–37.
- Carugo O (2007) Statistical validation of the root-mean-square-distance, a measure of protein structural proximity. *Protein Engineering Design and Selection* 20:33–37. <https://doi.org/10.1093/protein/gzl051>
- Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH (2015) Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. *Theoretical and Applied Genetics* 128:1647–1667. <https://doi.org/10.1007/s00122-015-2555-4>
- Combs GF (2012) *The Vitamins Fundamental Aspects in Nutrition and Health A Sustainability Challenge : Food Security for All*. 4th ed. Academic Press. New York.
- De Conto V, Braz ASK, Perahia D, Scott LPB (2015) Recovery of the wild type atomic flexibility in the HIV-1 protease double mutants. *Journal of Molecular Graphics and Modelling* 59:107–116. <https://doi.org/10.1016/j.jmgm.2015.04.006>
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15. Retrieved from <http://ci.nii.ac.jp/naid/20000864368>
- Fu Z, Yan J, Zheng Y, Warburton ML, Crouch JH, Li JS (2010) Nucleotide diversity and molecular evolution of the PSY1 gene in Zea mays compared to some other grass species. *Theoretical and Applied Genetics* 120:709–720. <https://doi.org/10.1007/s00122-009-1188-x>
- Gady ALF, Vriezen WH, Van de Wal MHB, Huang P, Bovy AG, Visser RGF, Bachem CWB (2012) Induced point mutations in the phytoene synthase 1 gene cause differences

- in carotenoid content during tomato fruit ripening. *Molecular Breeding* 29:801–812. <https://doi.org/10.1007/s11032-011-9591-9>
- Harjes CE et al (2008) *Science* 319:16–18.
- Howeler R, Lutaladio N, Thomas G (2013) *Save and Grow: Cassava. A Guide to Sustainable Production Intensification (First)*. Rome: Food and Agriculture Organization of the United Nations (FAO).
- Huang F, Nau WM (2003) A conformational flexibility scale for amino acids in peptides. *Angewandte Chemie - International Edition* 42:2269–2272. <https://doi.org/10.1002/anie.200250684>
- Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research* 9:868–877. <https://doi.org/10.1101/gr.9.9.868>
- Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14:33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
- Kempner ES (1993) Structural that control. *FEBS Letters* 326:4–10.
- Kermode AR (2005) Role of abscisic acid in seed dormancy. *Journal of Plant Growth Regulation* 24:319–344. <https://doi.org/10.1007/s00344-005-0110-2>
- Kim H, Schmidt CJ, Decker KS, Emara MG (2003) A double-screening method to identify reliable candidate non-synonymous SNPs from chicken EST data. *Animal Genetics* 34: 249–254. <https://doi.org/10.1046/j.1365-2052.2003.01003.x>
- Kingsley L and Lill M (2015) *Substrate Tunnels in Enzymes: Structure-Function Relationships and Computational Methodology*. *Proteins* 83:599–611
- Liu D, Chen W, Zhang B, Liu D, Liu B, Zhang H (2015) Diversity and distribution of puroindoline-D1 genes in *Aegilops tauschii*. *Genetic Resources and Crop Evolution* 63:1–13. <https://doi.org/10.1007/s10722-015-0271-1>
- Liu J, Tan H, Rost B (2002) Loopy proteins appear conserved in evolution. *Journal of Molecular Biology* 322:53–64. [https://doi.org/10.1016/S0022-2836\(02\)00736-2](https://doi.org/10.1016/S0022-2836(02)00736-2)
- López-Emparán A, Quezada-Martinez D, Zúñiga-Bustos M, Cifuentes V, Iñiguez-Luy F, Federico ML (2014) Functional analysis of the Brassica napus L. Phytoene Synthase (PSY) gene family. *PLoS ONE* 9:1–19. <https://doi.org/10.1371/journal.pone.0114878>
- Maass D, Arango J, Wüst F, Beyer P, Welsch R (2009) Carotenoid crystal formation in Arabidopsis and carrot roots caused by increased phytoene synthase protein levels. *PLoS ONE* 4:1-12. <https://doi.org/10.1371/journal.pone.0006373>
- Moise AR, Al-Babili S, Wurtzel ET (2014) Mechanistic aspects of carotenoid biosynthesis. *Chemical Reviews* 114:164–193. <https://doi.org/10.1021/cr400106y>
- Moose SP, Mumm RH (2008) *Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement*. *Plant Physiology* 147:969–977. <https://doi.org/10.1104/pp.108.118232>
- Nassar NMA, Junior OP, Sousa MV, Ortiz R (2009). Improving carotenoids and amino-acids in cassava. *Recent Patents on Food Nutrition Agriculture* 1:32–8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20653524>
- Nassar N, Vizzotto CS, Schwartz CA, Pires OR (2007) Cassava diversity in Brazil: The case of carotenoid-rich landraces. *Genetics and Molecular Research* 6:116–121.
- Niba LL, Bokanga MM, Jackson FL, Schlimme DS, Li BW (2002) Physicochemical Properties and Starch Granular Characteristics of Flour from Various *Manihot Esculenta* (Cassava) Genotypes. *Journal of Food Science* 67:1701–1705. <https://doi.org/10.1111/j.1365-2621.2002.tb08709.x>
- Nassar NM (2003) Cassava, *Manihot Esculenta* Crantz Genetic Resources: VI. Anatomy of a Diversity Center - Journals - NCBI. *Genet Mol Res* 2:214–222. Retrieved from <https://www.ncbi.nlm.nih.gov/labs/articles/14966687/>
- Owens BF et al (2014) A foundation for provitamin a biofortification of maize: Genome-wide

- association and genomic prediction models of carotenoid levels. *Genetics* 198:1699–1716. <https://doi.org/10.1534/genetics.114.169979>
- Paine JA et al (2005). Improving the nutritional value of Golden Rice through increased provitamin A content. *Nature Biotechnology* 23:482–487. <https://doi.org/10.1038/nbt1082>
- Pandit J et al (2000) Crystal Structure of Human Squalene Synthase: A KEY ENZYME IN CHOLESTEROL BIOSYNTHESIS. *Journal of Biological Chemistry* 275:30610–30617. <https://doi.org/10.1074/jbc.M004132200>
- Pauling L, Corey RB, Branson HR (1951) The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences* 37:205–211. <https://doi.org/10.1073/pnas.37.4.205>
- Pravda L et al (2014). Anatomy of enzyme channels. *BMC Bioinformatics* 15:379. <https://doi.org/10.1186/s12859-014-0379-x>
- R Core Team (2013) R: A Language and Environment for Statistical Computing
- Sali A, Blundell TL (1993) Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* 234:779–815. <https://doi.org/10.1006/jmbi.1993.1626>
- Salzburger W, Ewing GB, Von Haeseler A (2011) The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. *Molecular Ecology* 20:1952–1963. <https://doi.org/10.1111/j.1365-294X.2011.05066.x>
- Sánchez T et al (2014) Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and Hunter color techniques. *Food Chemistry* 151:444–451. <https://doi.org/10.1016/j.foodchem.2013.11.081>
- Shumskaya M, Bradbury LMT, Monaco RR, Wurtzel ET (2012) Plastid Localization of the Key Carotenoid Enzyme Phytoene Synthase Is Altered by Isozyme, Allelic Variation, and Activity. *The Plant Cell* 24:3725–3741. <https://doi.org/10.1105/tpc.112.104174>
- Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* 33:244–248. <https://doi.org/10.1093/nar/gki408>
- Stahl W, Sies H (2005) Bioactivity and protective effects of natural carotenoids. *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1740:101–107. <https://doi.org/10.1016/j.bbadis.2004.12.006>
- Stitzel NO, Binkowski TA, Tseng YY, Kasif S, Liang J (2004) topoSNP: a topographic database of non-synonymous single nucleotide polymorphisms with and without known disease association. *Nucleic Acids Research* 32:520D–522. <https://doi.org/10.1093/nar/gkh104>
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Tang J, Vosman B, Voorrips RE, van der Linden CG, Leunissen J a M (2006) QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. *BMC Bioinformatics* 7:1-15. <https://doi.org/10.1186/1471-2105-7-438>
- Umeno D, Tobias AV, Arnold FH (2002) Evolution of the C 30 Carotenoid Synthase CrtM for Function in a C 40 Pathway. *Journal of Bacteriology* 184:6690–6699. <https://doi.org/10.1128/JB.184.23.6690>
- Van den Burg B, Eijssink VG (2002) Selection of mutations for increased protein stability. *Current Opinion in Biotechnology* 13:333–337. [https://doi.org/10.1016/S0958-1669\(02\)00325-7](https://doi.org/10.1016/S0958-1669(02)00325-7)
- Welsch R et al (2010) Provitamin A Accumulation in Cassava (*Manihot esculenta*) Roots Driven by a Single Nucleotide Polymorphism in a Phytoene Synthase Gene. *The Plant*

- Cell 22:3348–3356. <https://doi.org/10.1105/tpc.110.077560>
- Yan BX, Sun QY (1997) Glycine residues provide flexibility for enzyme active sites. *Journal of Biological Chemistry* 272:3190–3194. <https://doi.org/10.1074/jbc.272.6.3190>
- Yan J, Kandianis CB, Harjes CE, Bai L, Kim E-H, Yang X, ... Rocheford T (2010) Rare genetic variation at *Zea mays crtRB1* increases β -carotene in maize grain. *Nature Genetics* 42:322–327. <https://doi.org/10.1038/ng.551>

3 ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA PARA CONTEÚDO DE AMIDO EM ACESSOS DE MANDIOCA DO BANCO DE GERMOPLASMA DA AMAZÔNIA ORIENTAL²

RESUMO

A mandioca (*Manihot esculenta* Crantz) é uma espécie que produz raízes amiláceas. Assim, realizou-se um estudo de associação genômica ampla (GWAS) com o objetivo de identificar marcadores SNPs e genes candidatos associados ao conteúdo de amido, para auxiliar o melhoramento genético sistemático da mandioca para níveis crescentes de amido nas raízes. Foi utilizada a plataforma de genotipagem por sequenciamento de alto rendimento para genotipar um painel de associação diversificada de 398 acessos e identificado 80.327 polimorfismos de nucleotídeo único (SNPs) distribuídos pelo genoma da mandioca. A fenotipagem foi realizada em 255 acessos. Para aplicação de GWAS foi utilizado o Modelo Linear Misto (MLM). Dentro de uma região genômica de aproximadamente 20kb do cromossomo 1, foi identificado um SNP associado ao conteúdo de amido na mandioca. Um levantamento da sequência do genoma da mandioca v6.1 posicionou esse SNP na região do locus Manes.01G048600, a anotação funcional deste gene relaciona a sua atividade biológica com a enzima glicose-6-fosfato 1-epimerase, a qual é ligada ao metabolismo de carboidrato.

Palavras-chave: Amido, Glicose-6-fosfato 1-epimerase, Gene, Metabolismo de carboidrato.

ABSTRACT

Cassava (*Manihot esculenta* Crantz) is a species that produces starchy roots. Thus, a genome-wide association study (GWAS) was conducted to identify SNPs and candidate genes associated with starch content, to support the systematic genetic improvement of cassava to increasing levels of starch in the roots. The high throughput genotyping platform was used to genotype a panel of diverse association of 398 accessions and identified 80,327 single nucleotide polymorphisms (SNPs) distributed by the cassava genome. A phenotype was performed in 255 accessions. For the GWAS application, the Mixed Linear Model (MLM). Within a genomic region of approximately 20kb of chromosome 1, a SNP associated with the starch content was identified in cassava. A survey of the cassava genome sequence v6.1 positioned this SNP in the locus Manes locus.01G048600, the functional annotation of this

² Este capítulo segue as normas da revista Crop Breeding and Applied Biotechnology

gene relates its biological activity to the enzyme glucose-6-phosphate 1-epimerase, which is linked to metabolism of carbohydrate.

Keywords: Carbohydrate metabolism, Gene, Glucose-6-phosphate 1-epimerase, Starch,

3.1 Introdução

A mandioca (*Manihot esculenta* Crantz) é uma espécie que produz raízes amiláceas e é considerada um alimento básico para mais de 800 milhões de pessoas em todo o mundo (Lebot 2009). É uma cultura alimentar cultivada extensivamente devido a sua tolerância à seca e produtividade estável, mesmo quando cultivada em solos de baixa fertilidade, e é a cultura tuberosa mais importante nas regiões tropicais da África, Ásia e América Latina, onde desempenha um papel importante para segurança alimentar em comunidades agrícolas mais pobres (Santisopasri *et al.* 2001; Monfreda *et al.* 2008; Salvador *et al.* 2014).

Na região norte do Brasil, a mandioca é cultivada em praticamente toda a sua extensão territorial, o que significa que o país tem variabilidade genética para adaptação a diferentes ambientes e condições climáticas (Moura *et al.* 2013). A mandioca é usada especialmente para o consumo humano, alimentação animal e na indústria, como a de celulose e têxtil. O amido é o carboidrato mais abundante da mandioca e é amplamente utilizado como substituto do amido de milho em alguns países (Bahnassey e Breene 1994), sendo é amplamente utilizado na indústria farmacêutica e alimentar por causa de suas propriedades únicas de espessamento, alta pureza, baixo custo e sua capacidade de formar pastas viscosas claras (Defloor *et al.* 1998; Chatakanonda *et al.* 2003), e no Brasil, a região Sul já vem se destacando pela forte aplicação da mandioca na indústria do amido (Dixon *et al.* 2003). Com a crescente demanda de amido, juntamente com o aumento significativo da população mundial, a maior parte das pesquisas focam na produção de cultivares com maior rendimento de amido por unidade de área cultivada (Saithong *et al.* 2013). Assim, o conteúdo de amido é um descritor quantitativo de imenso valor devido as demandas calóricas nas dietas das populações de regiões mais pobres e devido sua crescente aplicação industrial.

O Brasil apresenta uma grande diversidade de recursos genéticos de mandioca, considerada importante porque preserva a variabilidade genética das plantas de mandioca, que podem ser usadas no desenvolvimento de novas variedades de alta qualidade e de alto rendimento (Oliveira *et al.* 2014). A combinação da diversidade de recursos genéticos com os avanços recentes nas tecnologias de sequenciamento de nova geração (NGS) tornaram o estudo de associação genômico (GWAS) uma ferramenta favorável para mapeamento de caracteres complexos em plantas, a exemplo da mandioca (Davey *et al.* 2011). Atualmente,

tornou-se prático usar plataformas de genotipagem de alta densidade, como genotipagem por sequenciamento (GBS) para gerar rapidamente milhões de marcadores que são distribuídos por todo o genoma para facilitar o melhoramento genético de plantas (He *et al.* 2014).

Por meio do GWAS variações genéticas importantes podem ser exploradas e com isso melhorar características de importância agrícola e assim aumentar o fornecimento de alimentos biofortificados para população (Ceballos *et al.* 2015). Atualmente vários trabalhos que aplicam GWAS em diversas espécies vegetais vem sendo desenvolvidos no milho (Chen *et al.* 2015), no sorgo (Morris *et al.* 2013), no arroz (Spindel *et al.* 2016), na soja (Sonah *et al.* 2015) e eucalipto (Muller *et al.* 2017). Com mais de 97% do genoma de mandioca sequenciado (Prochnik *et al.*, 2012), agora é possível usar GBS (Elshire *et al.* 2011) e GWAS (Oliveira *et al.* 2012) para ancorar com precisão os polimorfismos de nucleotídeos únicos (SNPs) ligados a características de interesse. De fato, essas plataformas facilitam a inferência de funções de tais polimorfismos com base em anotações de regiões genômicas onde estão posicionadas. Por exemplo, Esuma *et al.* (2016) usaram SNP gerados a partir da plataforma GBS para realizar GWAS e identificar SNP importantes para a síntese de carotenóides na mandioca.

Por isso, foi realizado um teste GWAS com 80.327 marcadores SNP e 398 acessos de mandioca pertencente ao Banco ativo de Germoplasma (BAG) da Amazônia Oriental, com o objetivo de identificar marcadores SNPs e genes candidatos associados ao conteúdo de amido, para auxiliar o melhoramento genético sistemático da mandioca para níveis crescentes de amido nas raízes.

3.2 Materiais e métodos

3.2.1 Material vegetal

Foi utilizado no presente estudo um painel diverso de 398 acessos, incluindo variedades, introduções e cruzamentos, mantidos no banco ativo de germoplasma da Amazônia Oriental, em que 324 acessos são classificados como mandioca brava, 49 são mandioca mansa ou macaxeira e 25 são mandiocas açucaradas ou mandiocabas. Os acessos foram coletados nos estados do Pará, Amazonas, Rondônia, Amapá, Mato Grosso do Sul, Paraíba e Paraná, Brasil.

3.2.2 Fenotipagem

Foram fenotipados 255 acessos de mandioca, sendo 213 mandioca brava, 39 macaxeira e uma mandioca açucarada, por meio do descritor quantitativo conteúdo de amido na raiz de acessos de mandioca pertencente ao Banco Ativo de Germoplasma (BAG) da Amazônia Oriental.

O BAG está localizado no município de Belém, Pará, Brasil (01° 27' 21'' S e 48° 30' 16'' W) na sede da Embrapa Amazônia Oriental. A temperatura média anual da região é de 28,0°C, a umidade relativa do ar de 80% e a precipitação média anual de 2.800 mm (Instituto Nacional de Meteorologia - INMET, 2009). Os acessos estão dispostos no banco em fileiras simples compostas de nove plantas por acesso, com espaçamento de 1,0 m entre fileiras e 1,0 m entre plantas e são renovados quando atingem a idade de 12 a 14 meses. Os acessos não obedeceram a nenhum delineamento experimental específico.

As avaliações foram efetuadas no campo em plantas com 12 meses de idade. Os dados foram obtidos a partir de sete plantas úteis dentro de cada fileira, evitando-se as bordaduras. Após a coleta, as raízes podres foram descartadas e o conteúdo de amido da raiz foi mensurado por meio do método da balança hidrostática (Grossmann e Freitas 1950). Pesou-se uma amostra de 3 kg de raízes, lavadas e secas à sombra, e determinou-se o peso em água, em gramas, através de uma balança hidrostática. Determinou-se a porcentagem de amido na raiz pela fórmula: MS (Matéria seca) = 15,75 + 0,0564 R, em que R é o peso de 3 kg de raízes na balança hidrostática, em seguida a porcentagem de amido das raízes foi calculada, subtraindo-se do teor de matéria seca a constante 4,65.

3.2.3 Genotipagem

O DNA total foi extraído a partir de folhas jovens de acordo com a metodologia Doyle e Doyle (1990) e a qualidade do DNA para cada amostra foi cuidadosamente verificado por eletroforese em gel.

Foi realizada a genotipagem por sequenciamento (*Genotyping-by-sequencing* -GBS) dos 398 acessos no IDG (*Institute for Genomic Diversity*) da Universidade de Cornell nos Estados Unidos (EUA) gerando os marcadores moleculares do tipo SNP (Elshire *et al.* 2011). Na aplicação da técnica de GBS foi usada a enzima de restrição ApeKI, que é uma endonuclease de restrição do tipo II que reconhece uma sequência degenerada de 5 pb (GCWGC, onde W é A ou T), sendo utilizada para cortar regiões de DNA não repetitivo. A

enzima de restrição ApeKI foi selecionada porque apresenta uma sensibilidade parcial a metilação do DNA e corta o DNA genômico a baixa frequência, evitando assim as regiões genômicas repetitivas e é extremamente útil na promoção da ligação eficiente do adaptador com o DNA (Peterson *et al.* 2014).

As leituras das sequências dos diferentes acessos foram identificadas utilizando códigos de barras e alinhadas com a sequência do genoma de referência da mandioca v6.1 (Glaubitz *et al.* 2014) e com isso foi aplicado o chamado *Discovery Pipeline* para obtenção dos SNP, utilizando o *pipeline* TASSEL-GBS seguindo os procedimentos bioinformáticos descritos por Elshire *et al.* (2011) e Glaubitz *et al.* (2014).

Para garantir a alta qualidade das análises, os dados genotípicos foram processados usando o TASSEL v5.2.25 (Bradbury *et al.* 2007). Devido ao grande número de dados perdidos gerados pela técnica GBS, o conjunto de dados foi filtrado permitindo 0.8 de dados perdidos, o que exigiu a imputação dos marcadores não genotipados (Romay *et al.* 2013; Swarts *et al.* 2014). Anteriormente a imputação, foi realizada a remoção das inserções e deleções da sequência bruta e os dados foram filtrados por heterozigosidade, excluindo sítios com frequência maior ou igual a 0,8 de heterozigosidade e filtrado pela Menor frequência alélica do alelo alternativo (*Minor Allele frequency* ou MAF) igual a 0.01. Em seguida o conjunto de dados filtrado foi imputado usando o software Beagle v4.0 com base em um algoritmo que procura o vizinho mais próximo em pequenas janelas SNP em todo o genoma (Browning e Browning 2013; Swarts *et al.* 2014) e apresentou acurácia de 93 %. Após imputação os dados foram novamente filtrados por MAF 0.05.

Os dados filtrados apresentaram um conjunto de marcador reduzido compreendendo 80.327 SNP com uma MAF robusta desejável para estimar a estrutura da população e a matriz de parentesco, sendo ambos importantes para as subseqüentes análises estatísticas.

3.2.4 Análises Estatísticas

Realizou-se a análise de componentes principais (PCA), a partir dos SNP filtrados para averiguar o agrupamento entre os acessos, bem como verificar a possível existência de estrutura populacional entre os acessos. Para isso foi utilizado o software Tassel v5.2.25 (Bradbury *et al.* 2007), para gerar os componentes principais (CP) e em seguida os valores foram exportados para serem usados no ambiente do software R (R Development Core Team, 2011), sendo os dois primeiros componentes da PCA utilizados para desenhar um gráfico de dispersão para visualizar a diferenciação genética entre acessos.

Após as filtragens e imputação para remoção dos monomórficos e SNP de baixa qualidade, os 80.327 SNPs estavam disponíveis para o painel de associação de 398 indivíduos. Na execução de GWAS, foram considerados para as análises estatísticas os 398 acessos com dados genotípicos e fenotípicos, sendo que os acessos que não apresentaram a fenotipagem, foram considerados como dados perdidos (NA). Foi aplicado o Modelo Linear Misto (MLM), considerando a matriz de parentesco e os três primeiros componentes principais constituíram a estrutura da população. O software Tassel v5.2.25 (Bradbury *et al.* 2007) foi usado para aplicação do MLM e gerar os p-valores.

As avaliações do modelo de associação foram baseadas no gráfico quantil-quantil (Q-Q plot), que é um gráfico dos valores observados de $-\log_{10}P$ versus valores esperados de $-\log_{10}P$ sob a hipótese nula de que não há associação entre um SNP e o fenótipo. A abordagem usada para identificar sinais de GWA baseou-se inteiramente nos menores valores de P obtidos a partir do MLM, onde os SNPs foram classificados com base na ordem ascendente dos seus valores de P . Os sinais de associação dos cromossomos foram visualizados a partir de parcelas de Manhattan plot geradas usando o pacote *qqman* de software R (Turner 2014). Foi utilizada a sequência v6.1 do genoma de referência da mandioca usando phytozome v12.0 (Goodstein *et al.* 2012) para pesquisar genes modelo nas proximidades de SNPs selecionados. Intervalos de até 20.36 kb foram considerados ao procurar genes de percurso na vizinhança de SNPs com possíveis associações.

3.3 Resultados e discussão

A mandioca é uma cultura altamente heterozigótica e com genoma considerado grande (Oliveira *et al.* 2012). A versão 6.1 do conjunto principal do genoma da mandioca é de aproximadamente 582,25 Mb com um total de 33.033 loci contendo transcritos de codificação de proteínas dispostos em 18 cromossomos e 2001 scaffolds que não foram ancorados em cromossomos (Goodstein *et al.* 2012; Prochnik *et al.* 2012). Portanto, para obter uma cobertura razoável de um genoma tão grande através do GBS, é necessário um grande número de SNPs distribuídos ao longo do genoma (Deschamps *et al.* 2012). Neste estudo, foram utilizados um total de 670.835 marcadores SNPs distribuídos em 18 cromossomos e scaffolds de 398 acessos. No total, foram identificados 3 milhões de tags em todos os acessos de mandioca, sendo aproximadamente 5 tags por SNPs.

Após a filtragem para remoção de SNPs não informativos e imputação do subconjunto genotípico, do total de 670.835 SNPs restaram 80.327 SNPs que foram selecionados e utilizados para análises subsequentes, assim um total de 88% dos SNPs foram perdido após a

imputação e filtragens. Os dados imputados, portanto, foram usados para realizar a análise de GWA.

Os acessos de mandioca utilizados neste estudo foram provenientes de banco ativo de germoplasma da Embrapa Amazônia Oriental. Assim eles apresentaram um conjunto de acessos de diversas origens genéticas, o que seria uma possível fonte de estrutura populacional.

Visando descrever a estrutura da população para formação de um painel de associação para este estudo, foi utilizado PCA. Através dessa análise pode-se descrever o conjunto de dados constituído por 398 acessos caracterizados por 80.327 SNPs. Ao classificar os acessos em mandioca brava, macaxeira e mandioca açucarada, revelou-se um nível de diferenciação genética, com o grupo de mandioca açucarada derivando para a direita ao longo do segundo eixo do componente principal. Os acessos de mandiocas brava e macaxeiras tendenciaram a formação de grupos ao longo do primeiro e segundo eixo dos dois primeiros componentes principais, respectivamente, mas houve uma mistura de acessos, sem clara diferenciação entre eles. Portanto, o padrão de agrupamento sugeriu alguma estrutura dentro da população estudada (Figura 3.1). Esta aparente diferenciação genética forneceu a base para o uso de PCA e da matriz de parentesco no MLM para análise de associação, o que foi uma estratégia importante para minimizar quaisquer falsos positivos no resultado do mapeamento.

Por meio da distribuição da frequência alélica do alelo alternativo (MAF) dos dados não imputado, pode-se observar que grande parte dos SNPs apresentaram MAF menor que 5% (Figura 3.2). De acordo com Glaubitz *et al.* (2014) marcadores com MAF muito baixa devem ser eliminados, pois podem representar falsos positivos, devido erros de sequenciamento. Assim os dados foram filtrados com MAF de 5% para serem utilizados no MLM da análise de GWA.

A média para a característica conteúdo de amido para 255 acessos fenotipados foi de 27.02 %, com um intervalo de 19.00 a 34.79% (base úmida), apresentando distribuição normal (Figura 3.3), sendo semelhantes aos resultados encontrados por Gu *et al.* (2013).

Com base na análise de MLM para a característica conteúdo de amido, verificou-se que os valores observados perfazem uma linha abaixo dos esperados (Figura 3.4), com isso foram selecionados para pesquisa de anotações dos SNP os 10 menos p-valores (Tabela 3.1) e uma análise cromossômica mapeou o sinal para os cromossomos 1, 3, 8, 10, 11, e 14 do genoma da mandioca com 10 SNPs possivelmente associados a característica conteúdo de amido na mandioca (Figura 3.5), entre os quais o SNP S1_9830874, localizado no cromossomo 1, foi o mais associado ($P = 1.2622e-04$). Esuma *et al.* (2016) estudando GWAS

para conteúdo de carotenóides na raiz de mandioca utilizaram 179.310 marcadores e relatam que o baixo número de SNPs podem representar baixa cobertura do genoma e, assim, reduziram as chances de capturar mais polimorfismos com significância estatística. Apesar da natureza quantitativa do conteúdo de amido, que é conhecido por ser controlada por um grande número de genes com pequenos efeitos Liu *et al.* (2016) encontraram resultados semelhantes ao observado neste trabalho ao estudar GWAS na cultura do milho e identificaram um total de quatro SNPs associados com conteúdo de amido detectados nos cromossomos 1, 2 e 5, sendo que o mais significativo apresentou p-valor igual a $P = 2,9 \times 10^{-5}$.

Foi utilizada a sequência referência v6.1 do portal de Genômica Comparativa de Plantas, Phytozome v12.02 (Goodstein *et al.* 2012) para identificar genes anotados a uma distância de ± 20 Kb da região genômica ocupada pelos SNPs selecionados. Encontrou-se o gene Manes.01G048600, que é um gene de glucose-6-phosphate-1-epimerase conhecido por agir sobre carboidratos e derivados, sendo encontrado dentro da região genômica ocupada por SNPs significativos. O tamanho deste gene é de 2,89 kb que se estende de 9830200 a 9833091 pb no cromossomo 1. A anotação funcional deste gene relaciona a sua atividade biológica com a enzima glucose-6-phosphate 1-epimerase, a qual é fortemente ligada ao metabolismo de carboidrato.

Portanto o SNP S1_9830874 pode ser considerado importante para a via metabólica de carboidratos, pois este polimorfismo foi localizado na região de um gene YMR099C que codifica a proteína glucose-6-phosphate-1-epimerase previamente caracterizado com funções ligadas ao metabolismo de carboidrato. Esta proteína presumivelmente catalisa a interconversão de alfa-D-glucose-6-fosfato para Beta-D-glicose-6-fosfato e acelera a velocidade da reação de duas a três vezes mais, além disso a glucose-6-phosphate-1-epimerase catalisa a reação de pelo menos mais dois açúcares de hexose-6-fosfato (Gal6P e Man6P) e portanto, pode apresenta atividade de mutarotase (Graille *et al.* 2006).

Muitas enzimas envolvidas no metabolismo de carboidratos são específicas para α ou β anômero do seu substrato, com isso o equilíbrio da anomerização do carboidrato desempenha um papel nos fluxos metabólicos gerais da célula. Por exemplo a glicose desidrogenase utiliza apenas β -D-glucose como substrato (Pauly e Pfeleiderer 1975). Assim a necessidade de uma enzima que interconverta os anômeros alfa e beta Glc6P é ainda justificada pela especificidade de diferentes enzimas envolvidas nas vias de metabolismo do carboidrato (Graille *et al.* 2006).

A atividade da enzima glucose-6-phosphate-1-epimerase na interconversão de alfa-D-glucose-6-fosfato para Beta-D-glucose-6-fosfato é importante para o processo de respiração celular, pois atua na glicólise e gliconeogênese. Yang *et al.* (2011) estudando diferenças na expressão gênica de três tipos de raízes de mandioca, a fim de obter uma melhor compreensão do desenvolvimento da raiz da mandioca no nível transcricional, descobriram que a glicólise/gliconeogênese era o processo mais importante entre 25 vias em que foram observadas alterações significativas na expressão gênica durante o desenvolvimento radicular da mandioca, e que a redução da atividade de algumas enzimas, poderia fazer com que a maior parte da α -D-glucose-6-fosfato produzida fosse transportada para o amiloplasto para participar do metabolismo do amido e sacarose. Portanto as propriedades físico-químicas associadas à qualidade e conteúdo de amido na mandioca estão correlacionadas com genes envolvidos na biossíntese de amido em fases distintas e estádios de desenvolvimento radicular (Saithong *et al.* 2013).

REFERÊNCIAS

- Bahnassey YA and Breene WM (1994) Rapid Visco-Analyzer (RVA) Pasting Profiles of Wheat, Corn, Waxy Corn, Tapioca and Amaranth Starches (*A. hypochondriacus* and *A. cruentus*) in the Presence of Konjac Flour, Gellan, Guar, Xanthan and Locust Bean Gums. **Starch - Stärke** **46**:134–141.
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. **Bioinformatics** **23**:2633–2635.
- Browning BL and Browning SR (2013) Improving the accuracy and efficiency of identity-by-descent detection in population data. **Genetics** **194**:459–471.
- Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH (2015) Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. **Theoretical and Applied Genetics** **128**:1647–1667.
- Chatakanonda P, Chatakanonda P, Chinachoti P, Sriroth K, Piyachomkwanb K, Chotineeranatb S, Tangc H, Hills B (2003) The influence of time and conditions of harvest on the functional behaviour of cassava starch - A proton NMR relaxation study. **Carbohydrate Polymers** **53**:233–240.
- Chen G, Chen G, Wang X, Hao J, Yan J, Ding J (2015) Genome-wide association implicates candidate genes conferring resistance to maize rough dwarf disease in maize. **PLoS ONE** **10**:1–14.

- Davey JW Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics** **12**:499–510.
- Defloor I, Dehing I and Delcour JA (1998) Physico-Chemical Properties of Cassava Starch. **Starch - Stärke** **50**:58–64.
- Deschamps S, Llaca V and May GD (2012) Genotyping-by-Sequencing in Plants, **Biology** **1**:460–483.
- Dixon AGO, Bandyopadhyay R, Coyne D, Ferguson R, Ferris SB, Hanna R, Hughes J, Ingelbrecht I, Legg J, Mahungu N, Manyong V, Mowbray D, Neuenschwander P, Whyte J, Hartmann P, Ortiz R (2003) Cassava: From poor farmers crop to pacesetter of African rural development. **Chronica Horticulturae** **43**:8–15.
- Doyle JJ and Doyle JL (1990) Isolation of plant DNA from fresh tissue. **Focus** **12**:13–15.
- Elshire RJ *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS ONE** **6**:1–10.
- Esuma W, Herselman L, Labuschagne MT, Ramu P, Lu F, Baguma Y, Bucker ES, Kawuki RS (2016) Genome-wide association mapping of provitamin A carotenoid content in cassava. **Euphytica** **212**:97–110
- Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. **PLoS ONE** **9**:1–11.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: A comparative platform for green plant genomics. **Nucleic Acids Research** **40**:1178–1186.
- Graille M, Baltaze JP, Leulliot N, Liger D, Quevillon-Cheruel S, Van Tilbeurgh H (2006) Structure-based functional annotation: Yeast ymr099c codes for a D-hexose-6-phosphate mutarotase. **Journal of Biological Chemistry** **281**:30175–30185.
- Grossmann J e Freitas AC (1950) Determinação do teor de matéria seca pelo peso específico em raízes de mandioca. **Revista Agrônômica** **14**:75–80.
- Gu B, Yao Q, Li K, Chen S (2013) Change in physicochemical traits of cassava roots and starches associated with genotypes and environmental factors. **Starch/Staerke** **65**:253–263.
- He J, Zhao X, Laroche A, Lu Z-X, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS) an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. **Frontiers in Plant Science** **5**:1–9.
- Lebot V (2009) **Tropical root and tuber crops : cassava sweet potato, yams and aroids.**

First. Wallingford UK: CABI.

Liu N, Xue Y, Guo Z, Li W, Tang J (2016) Genome-Wide Association Study Identifies Candidate Genes for Starch Content Regulation in Maize Kernels. **Frontiers in Plant Science** 7:1–8.

Monfreda C, Ramankutty N and Foley JA (2008) Farming the planet: 2. Geographic distribution of crop areas yields, physiological types, and net primary production in the year 2000. **Global Biogeochemical Cycles** 22:1–19.

Morris GP, Ramu P, Deshpande SP, Hashc CT, Shahb T, Upadhyayab HD, Riera-Lizarazub O, Brownd PJ, Acharyae CB, Mitchelle SE, Harrimane J, Glaubitze JC, Bucklere ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. **Proceedings of the National Academy of Sciences** 110:453–458.

Moura EF, Tomé J, Neto DF, Sampaio JE, Tuloza D, Silva DA (2013) Identification of duplicates of cassava accessions sampled on the North Region of Brazil using microsatellite markers. **Acta Amazonica** 43:461–468.

Müller BSF, Neves LG, Filho JEA, Resende Jr MFR, Muñoz PR, Santos PET, Filho EP, Kirst M, Grattapaglia D (2017) Genomic Prediction in Contrast to a Genome-Wide Association Study in Explaining Heritable Variation of Complex Growth Traits in Breeding Populations of Eucalyptus. **BMC Genomics** 18:1–17. doi:10.1186/s12864-017-3920-2.

Oliveira EJ, de Resende MDV, Santos SV, Ferreira FF, Oliveira GAF, Silva MS, Oliveira LA, Aguilar-Validoso CI (2012) Genome-wide selection in cassava. **Euphytica** 187:263–276.

Oliveira EJ, Ferreira CF, da Silva Santos V, de Jesus ON, Oliveira GAF, da Silva MS (2014) Potential of SNP markers for the characterization of Brazilian cassava germplasm. **Theoretical and Applied Genetics** 127:1423–1440.

Pauly HE and Pfeleiderer G (1975) D-glucose dehydrogenase from *Bacillus megaterium* M 1286: purification properties and structure. **Hoppe-Seylers Zeitschrift fur physiologische Chemie** 356:1613–1623.

Peterson GW, Dong Y, Horbach C, Fu YB (2014) Genotyping-by-sequencing for plant genetic diversity analysis: A lab guide for SNP genotyping. **Diversity** 6:665–680.

Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar S, Rounsley S (2012) The Cassava Genome: Current Progress Future Directions. **Tropical Plant Biology** 5:88–94.

R Development Core Team (2011) R: a language and environment for statistical computing, 3.2.1 Document freely available on the internet at: <http://www.r-project.org>. Viena, Austria: R Foundation for Statistical Computing.

- Romay MC, Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, McMullen MD, Holland JB, Buckler ES, Gardner CA (2013) Comprehensive genotyping of the USA national maize inbred seed bank. **Genome Biology** **14**:R55.
- Saithong T, Rongsirikul O, Kalapanulak S, Chiewchankaset P, Siriwat W, Netrphan S, Suksangpanomrung M, Meechai A, Cheevadhanarak S (2013) Starch biosynthesis in cassava: a genome-based pathway reconstruction and its exploitation in data integration. **BMC Systems Biology** **7**:7–75.
- Salvador EM, Steenkamp V and McCrindle CME (2014) Production, consumption and nutritional value of cassava (*Manihot esculenta*, Crantz) in Mozambique: An overview. **Journal of Agricultural Biotechnology and Sustainable Development** **6**:29–38.
- Santisopasri V, Kurotjanawong K, Chotineerarat S, Piyachomkwan K, Sriroth K, Oates CG (2001) Impact of water stress on yield and quality of cassava starch. **Industrial Crops and Products** **13**:115–129.
- Sonah H, O'Donoghue L, Cober E, Rajcan I, Belzile F (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. **Plant Biotechnology Journal** **13**:211–221.
- Spindel JE, Begum H, Akdemir D, Collard B, Redoña E, Jannink J-L, McCouch S (2016) Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. **Heredity** **116**:395–408.
- Swarts K, Li H, Navarro JAR, An D, Romay MC, Hearne S, Acharya C, Glaubitz JC, Mitchell S, Elshire RJ, Buckler ES, Bradbury PJ (2014) Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. **The Plant Genome** **7**.
- Turner SD (2014) qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. <http://dx.doi.org/10.1101/005165>. Accessed on 12 June, 2017
- Yang J An D and Zhang P (2011) Expression Profiling of Cassava Storage Roots Reveals an Active Process of Glycolysis/Gluconeogenesis. **Journal of Integrative Plant Biology** **53**:193–211.

Tabela 3.1 Classificação ascendente dos 10 menores P-valores para o fenótipo conteúdo de amido da raiz de mandioca, gerados a partir do MLM e utilizados na análise de anotações dos SNP.

Fenótipo	SNP	Cromossomo	P-valor
Conteúdo de amido	S10_3680246	10	4,5806E-05
Conteúdo de amido	S8_31528500	8	5,1295E-05
Conteúdo de amido	S3_29049565	3	5,4488E-05
Conteúdo de amido	S1_10578283	1	6,6691E-05
Conteúdo de amido	S1_20716113	1	9,2854E-05
Conteúdo de amido	S1_9830874	1	1,2622E-04
Conteúdo de amido	S14_525178	14	1,6300E-04
Conteúdo de amido	S11_10636967	11	1,7447E-04
Conteúdo de amido	S10_5590819	10	1,7852E-04
Conteúdo de amido	S10_5590806	10	1,8788E-04

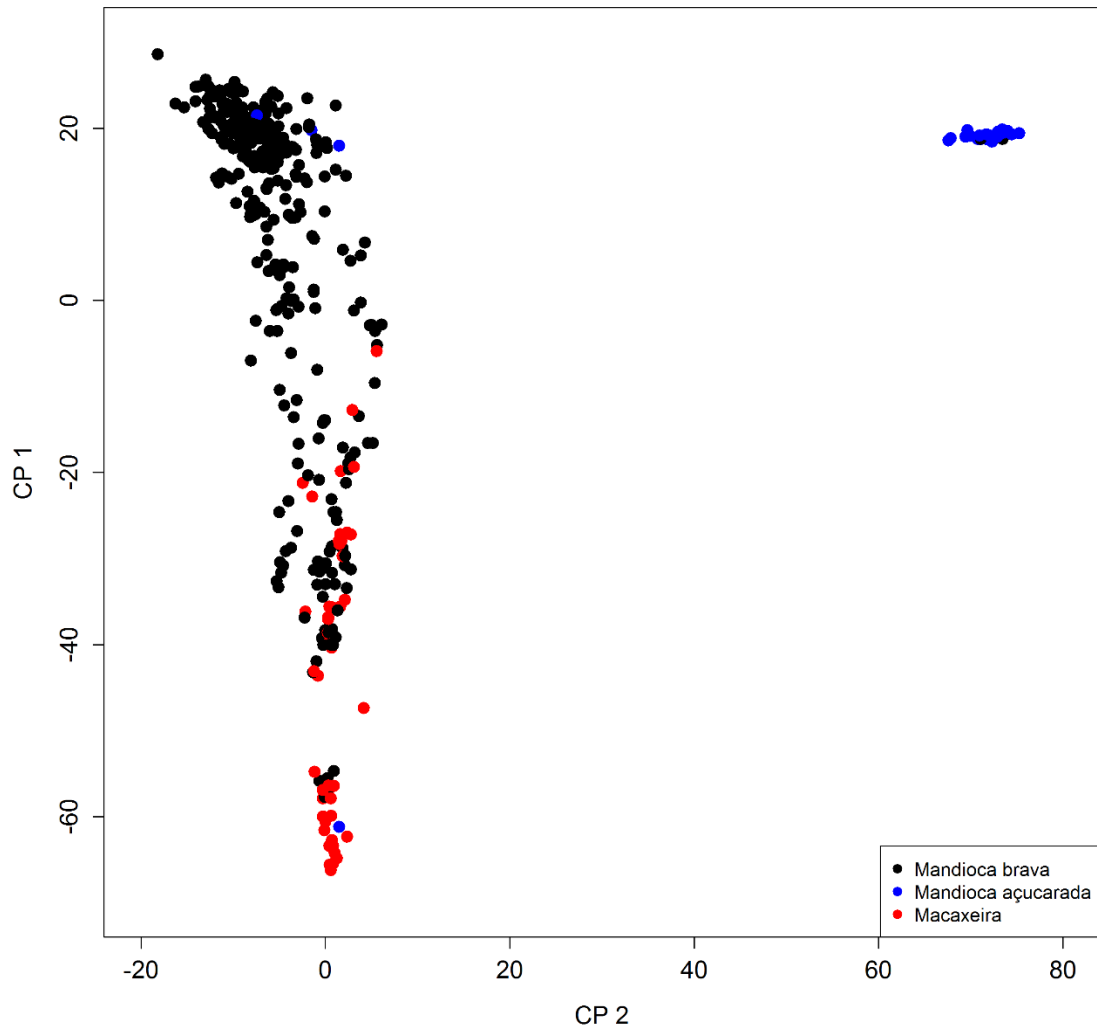


Figura 3.1. Dispersão de 398 acessos com base na análise de componentes principais. CP 1: componente principal 1, CP 2: componente principal 2.

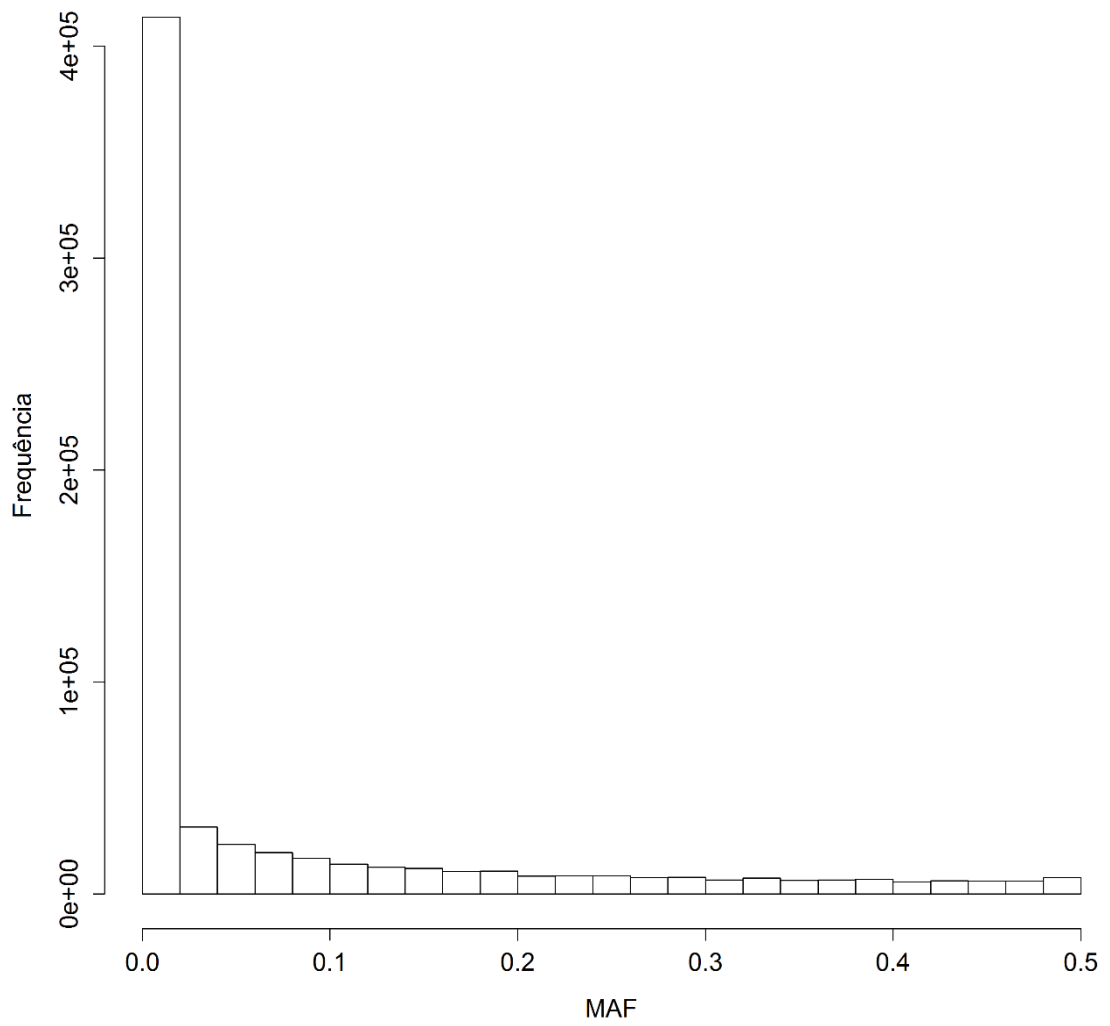


Figura 3.2. Distribuição da frequência do alelo alternativo (MAF- *minor allele frequency*), com base nos dados de genotipagem não filtrados.

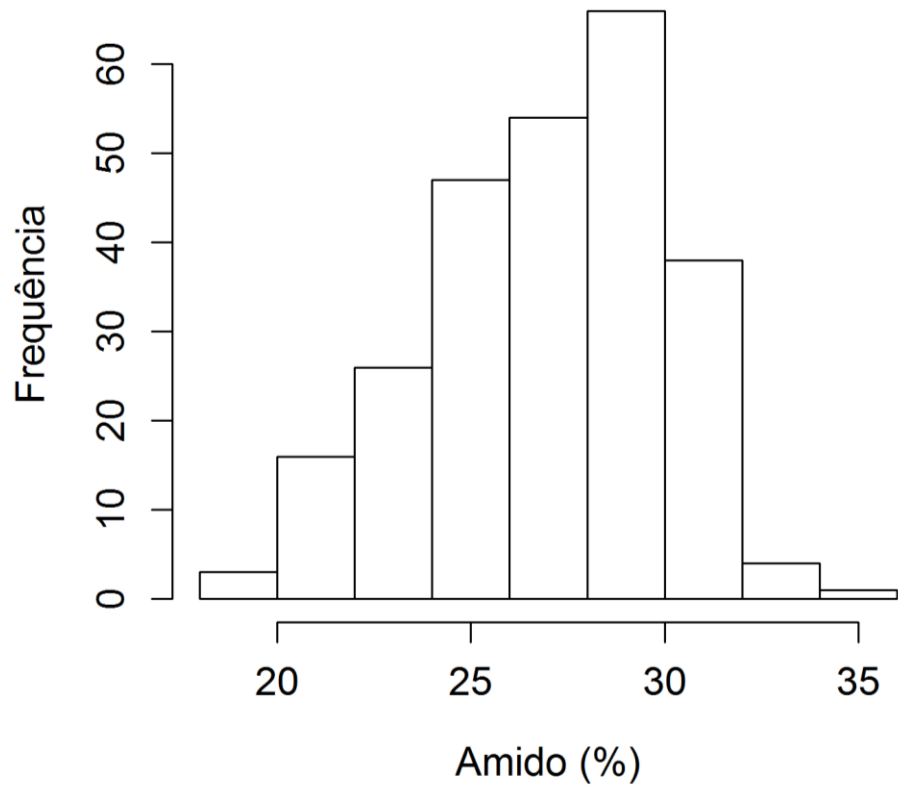


Figura 3.3. Distribuição da frequência do conteúdo de amido na raiz de mandioca.

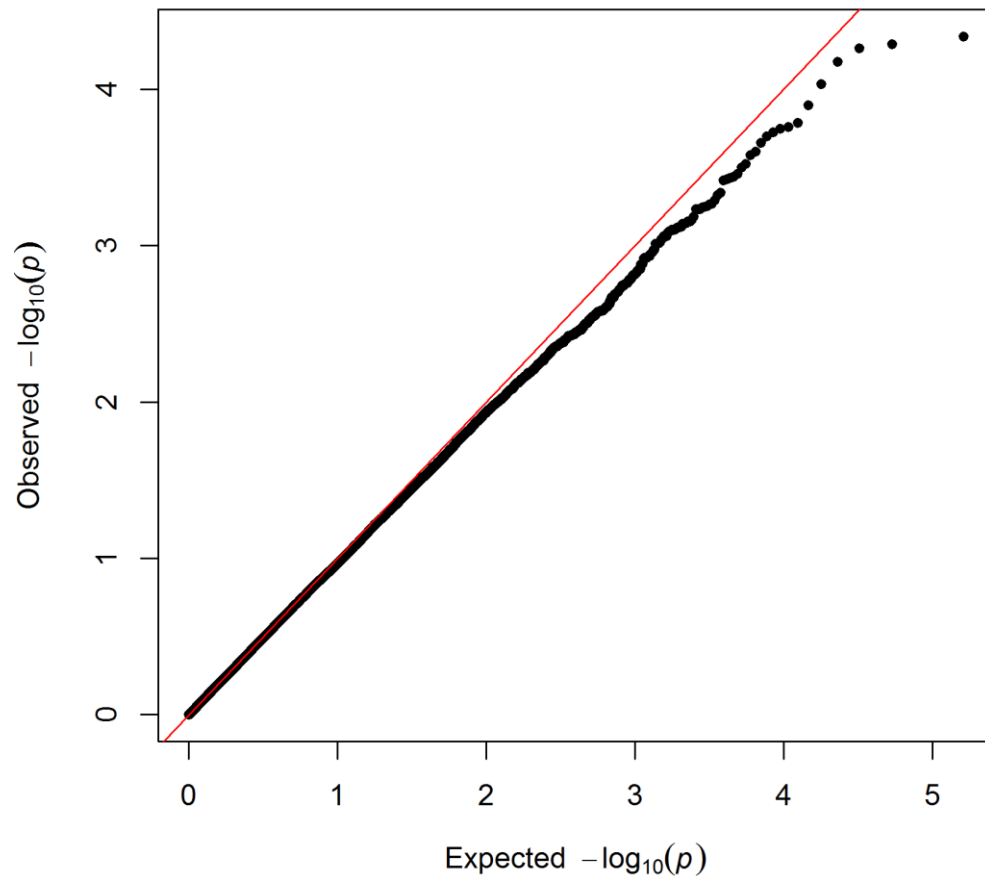


Figura 3.4. Gráfico quantil-quantil dos valores de $-\log_{10}(p)$ esperado contra o observado para a característica percentual de amido.

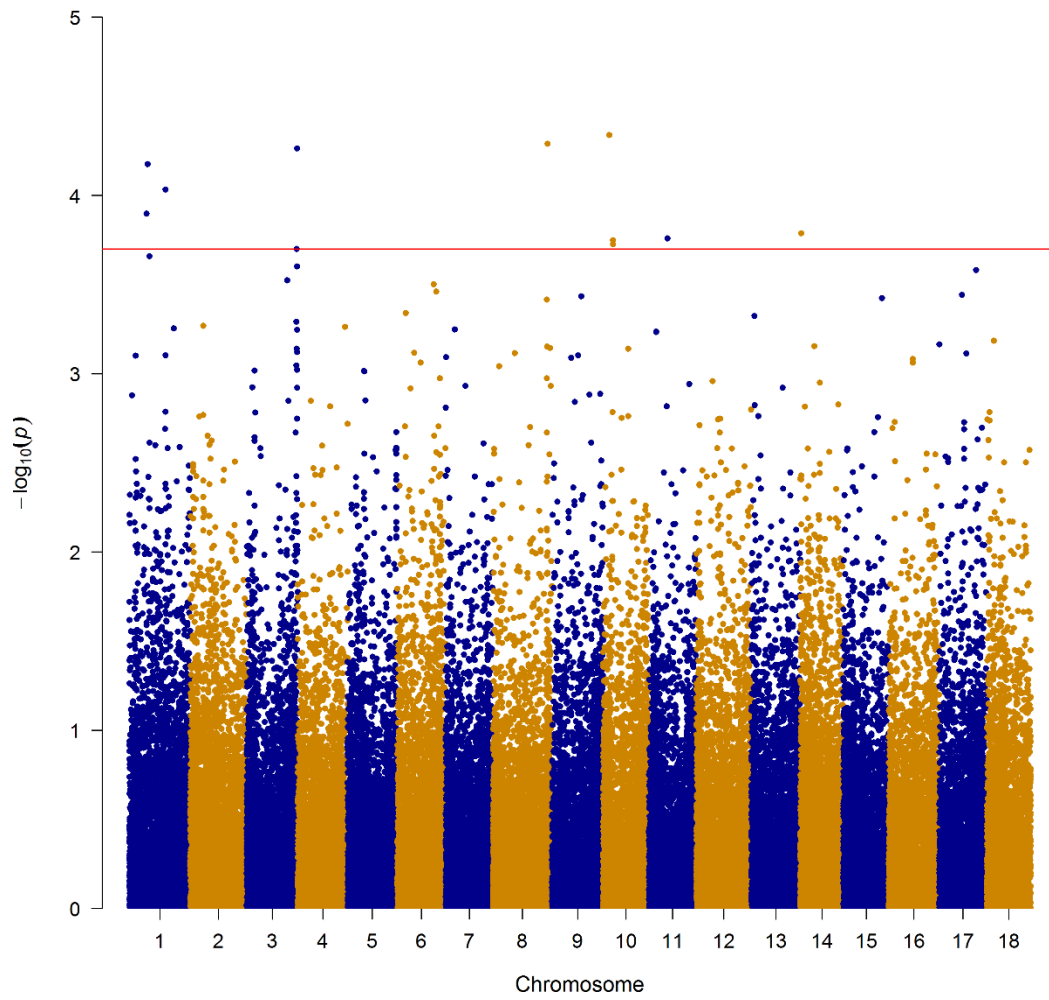


Figura 3.5. Manhattan plot da análise de associação genômica entre $-\log_{10}(p)$ (p-valor) e o loci do SNP em ordem sequencial por cromossomo, para característica conteúdo de amido.

4 CHEMICAL ROOT TRAITS DIFFERENTIATE GENOTYPES OF ‘BITTER’ AND ‘SWEET’ CASSAVA (*Manihot esculenta*) SAMPLED IN THE AMAZON³

ABSTRACT

‘Bitter’ and ‘sweet’ cassava are normally distinguished based on HCN content on their roots. Besides, Brazilian farmers tend to select ‘sweet’ cassava based on taste and cooking aspects. The aim of this study was to characterize chemical traits of roots of ‘bitter’ and ‘sweet’ cassava sampled in the Amazon and to verify genetic relations among accessions based on these traits. It was evaluated: moisture, ashes, total soluble solids, total titrateable acidity, pH, total carotenoids, free and total cyanide, crude protein, glucose, fructose, sucrose and starch. Considerable phenotypic variation was detected for the traits. Besides free and total cyanide, traits like sugars and moisture had different averages for ‘bitter’ and ‘sweet’ cassavas. Starch and free and total cyanide concentrations were the least important traits to discriminate genotypes. ‘Bitter’ and ‘sweet’ cassavas were grouped in different clusters in the dendrogram performed by UPGMA method and in the dispersion of principal component analysis.

Key words: carotenoids, cyanide, multivariate analysis, starch.

4.1 Introduction

Cassava (*Manihot esculenta* Crantz) is an important source of calories for the world’s population and it is widely grown in tropical and subtropical regions (Lebot 2009), due to its capacity to adapt to unfavorable conditions. Nigeria, Thailand, Indonesia, Brazil and the Democratic Republic of the Congo are the main world producers of roots (FAO 2014). Recent studies have showed that cassava was domesticated in south-western Amazonia rim (Léotard *et al.* 2009) and it is believed that Amazonia is an important source of genetic resources of cassava.

In countries where cassava is highly consumed, the development of nutrient-rich genotypes would be advantageous to prevent diseases and improve diets. The germplasm characterization of accessions maintained in germplasm banks is important to support genetic breeding programs, in order to identify promising genotypes that can generate biofortified cultivars (Carvalho *et al.* 2011, Sánchez *et al.* 2014). Besides, the quality of food products derived from cassava roots depend on its chemical composition. However, few studies aim the

³ Este capítulo segue as normas da revista Crop Breeding and Applied Biotechnology

chemical characterization of cassava roots that are traditionally grown or maintained in germplasm banks in the Amazon region.

The germplasm bank of cassava from Eastern Amazon maintains different genotypes sampled in the Brazilian Amazon, mostly from traditional farmers. In the Amazon, farmers tend to separate ‘bitter’ from ‘sweet’ cassavas. Currently, the landraces of cassava are classified in two types: ‘sweet’ cassava, also known as ‘aipim’, ‘macaxera’, table cassava and ‘bitter’ cassava, also known as cassava for industry (McKey *et al.* 2010). This classification is related to the capacity of cyanide (HCN) liberation, a highly toxic substance if ingested. Sweet cassavas has less than 100 mg of cyanogenic compounds per kg⁻¹ of fresh root, while ‘bitter’ cassavas has more than 100 mg of cyanogenic compounds (McKey *et al.* 2010). Thus, the aim of this study was to evaluate chemical traits of roots of ‘bitter’ and ‘sweet’ cassavas of the germplasm bank of Eastern Amazon.

4.2 Material and methods

It was analyzed 46 genotypes of cassava sampled in different cities of the States of Pará, Amazonas, Amapá and one from Bahia, classified as ‘bitter’ or ‘sweet’ cassava (Appendix Table 4.1) and maintained in a germplasm bank. The genotypes were cultivated in the city of Igarapé Açu, State of Pará, Brazil (lat 01° 07’ S, long 47° 36’ W and alt 54 m asl). The climatic conditions of this region are of Am type according to Koppen’s classification. Its average rain precipitation is of 2495 mm and average temperature per year is 26.4 °C.

The planting was done on a 1.0 m x 1.0 m spacing, with tillage and plowing for soil preparation and only one fertilizer application performed with nitrogen, phosphorous and potassium, using 10 - 28 - 20 formula 35 days after the planting of the stakes, where 40 g of fertilizer were used per planting spot. No irrigation was performed.

The harvest was carried out one year later. It was sampled three plants per genotype. Thus, chemical analysis was realized using three repetitions per genotype, where each repetition was represented by two roots per plant. Roots were mixed to compose a repetition. The roots were washed with water, packed in plastic bags and stored in a cold chamber at -18 a -20 °C for conservation until analyses. Roots were analyzed for 13 chemical traits. The analysis of moisture, gravimetric method n° 920.151, ashes by gravimetric method n° 940.26, pH by the method n° 981.12, total titratable acidity (TTA) by the method n° 942.15B and total soluble solids (TSS) by the method n° 932.12 were performed according to AOAC (1997). Crude protein concentration was established using the micro-Kjeldahl technique, based on

hydrolysis and later distillation of the sample using 6.25 x % N factor, in accordance with the AOAC (1997) method.

Determination of cyanogenic compounds (free and total cyanide) was obtained in accordance with Essers *et al.* (1993). For release of glycoside cyanide, the linamarin enzyme was used, which was extracted from the cassava root pulp (Essers *et al.* 1993).

The analysis of total carotenoids (TC) was performed in accordance with the methodology described by Rodriguez-Amaya (2001). Reading for determination of absorbency and quantification of glucose, fructose and sucrose sugar was done on spectrophotometer fit for Elisa plates on wave length of 340 nm, according to Stitt *et al.* (1989). Starch content in the roots was determined according to the method of hydrostatic balance (Grossman and Freitas 1950).

Data was submitted to analyses of descriptive statistics, where minimum, maximum and medium values and variation amplitude, coefficient of experimental variation and normality test were estimated to describe the variability of chemical characters. Analyses of phenotypic correlation of Spearman at 1% probability were performed for the traits. It was used the R 3.0.3 software package. To calculate the normality test of Rayan-Joiner (similar to Saphiro-Wilk's test) at 1%, it was used the Minitab 16 software (Minitab Inc 2008).

Principal components and clustering methods were used considering the means of the 13 traits for each genotype. To calculate principal components (PCs), it was used the software Minitab 16. The relative importance of traits was assessed at the discretion of the weight variables in eigenvectors. The eigenvectors and eigenvalues were obtained from the correlation matrix of standardized data of original values. Then, the dissimilarities between genotypes were visualized by the means of dispersion plotted in a two-dimensional graphic. Data were standardized to calculate the mean Euclidean distances among the 46 genotypes. Based on the matrix of genetic dissimilarities, a dendrogram was generated using the unweighted pair group method with arithmetic mean (UPGMA) method. It was used the software Genes 5.1 (Cruz 2006). Groups of genotypes in the dendrogram were determined based on 'NbClust' package of software R 3.0.3 (R Core Team) using Pseudot2 (Charrad *et al.* 2011). The adjustment of the matrix of dissimilarities was estimated according to the cophenetic correlation according to Sokal and Rohlf (1962) using Genes 5.1.

4.3 Results and discussion

Amplitude values were higher than means for most of the chemical traits, except for ashes, pH, moisture and starch (Table 4.1). Coefficients of experimental variation varied from

3% for pH to 93% for total cyanide (Table 4.1) and high values were detected for almost all characters, with the exception of TTA, TSS, ashes, pH, moisture and starch. Perhaps the high values of coefficient of variation were due to analysis of accessions under the condition of germplasm bank, where no experimental design was used. The traits TTA, TSS, ashes, pH, moisture and starch showed normal distribution since the Rayan-Joiner test was non-significative (Table 4.1).

In general, there was a great variation for all traits evaluated (Table 4.2 and 4.3). The genotypes CPATU 516, CPATU 475 and CPATU 375 showed cyanide contents among 50 to 100 mg HCN kg⁻¹ (Table 4.2), but these genotypes were not identified by local farmers as 'sweet' cassavas. It has been detected that cyanide contents in a genotype can be influenced by local conditions (Burns *et al.* 2012, Gu *et al.* 2013). Besides the low cyanide contents detected for genotypes CPATU 475 and CPATU 516, these genotypes also showed moderate levels of total carotenoids (Table 4.3), and they can be used in genetic breeding programs to generate biofortified genotypes.

For the measurement of total carotenoids, it was analyzed 20 genotypes of 'bitter' cassava and 15 genotypes of 'sweet' cassava, after observation of the color of the root pulp for each genotype. The highest contents of total carotenoids detected in cassava genotypes were 18.41 µg g⁻¹, 14.09 µg g⁻¹, 11.34 µg g⁻¹, 11.24 µg g⁻¹ and 10.88 µg g⁻¹, all in accessions of 'bitter' cassavas (Table 4.3). Thus, these genotypes can be considered potential genitors for genetic breeding programs. 'Sweet' cassavas had lower total carotenoid contents, that varied from 2.08 µg g⁻¹ to 4.38 µg g⁻¹ (Table 4.3). Using direct crossings, it was obtained total carotenoids contents of 25.5 µg g⁻¹ in the cassava root of improved clones, which confirmed that classical genetic breeding can gradually increase the contents of carotenoids in the roots (Sánchez *et al.* 2014). All genotypes used in this study were landraces grown by traditional farmers in the Amazon and represent great potential for genetic breeding to improve carotenoids in cassava roots.

Protein contents in 'bitter' cassavas varied from 0.46 to 2.27% (Table 4.3) and for 'sweet' cassavas, they varied from 0.18 to 1.10% (Table 4.3). This variation was below the variation described by Ceballos *et al.* (2006) for protein contents, where they evaluated accessions from different countries in Asia and America and identified values that ranged from 0.95 to 6.42%.

'Bitter' cassavas presented higher values of starch and sugars than the group of 'sweet' cassava (Table 4.3). Starch contents were similar to values identified in another study with cassava genotypes sampled in Brazil (Vieira *et al.* 2015, Oliveira *et al.* 2016).

Spearman correlation values were calculated for the 13 chemical traits (Table 4.4). TTA was negatively correlated with free cyanide and starch (Table 4.4). Negative correlation of glucose and sucrose with pH confirmed results obtained by Feltran *et al.* (2004), which concluded that higher levels of pH leads to low accumulation of sugars. Negative correlation between free cyanide and moisture indicates that the increase in water contents promotes fast solubilization and volatilization of cyanide and higher amounts of water in the root may reduce cyanide concentration as well. Negative correlation among cyanide and moisture may be interesting to explore in the manufacturing of products derived from cassava root. Starch was positively correlated with total cyanide (Table 4.4). The group of ‘sweet’ cassavas had lower contents of starch in their roots compared to the group of ‘bitter’ cassava (Table 4.3). It needs to be further investigated if farmer’s selection of genotypes to be consumed as ‘macaxeras’ has influenced starch contents in the roots.

A principal component analysis was used to evaluate the most important descriptors to discriminate between ‘sweet’ and ‘bitter’ cassava. The first four principal components (PCs) explained 37%, 15%, 10% and 9% of the total variation, respectively. The trait with the highest association (eigenvectors) with the first PC is the main responsible for phenotypic variation and the trait with the highest association with the last PC is the least responsible (Jolliffe 1973). The first two PCs accounted for 52% of the variation, and sucrose and fructose showed the highest association with the first and the second components, eigenvectors 0.38 and -0.51, respectively. Ashes and total titratable acidity (TAA) were highly correlated with the third component, (-0.53 e -0.52), respectively, and total carotenoids (TC) was highly correlated with the fourth component (0.81) (Table 4.5).

The graphical dispersion of the 46 genotypes based on the first two principal components (Figure 4.1) showed that the chemical traits were able to separate ‘bitter’ from ‘sweet’ genotypes. Genotypes of ‘bitter’ cassava were dispersed over three quarters of the graph. Three genotypes were dispersed separately at the top of the graph: CPATU 475, CPATU 465 and CPATU 466, and they showed lower contents of TC and high contents of glucose and fructose. Padonou *et al.* (2005) found similar results, where total sugars helped to discriminate among genotypes of ‘bitter’ cassava. Genotypes of ‘sweet’ cassava were allocated in the left-hand quarter. The genotype CPATU 283 (S) was plotted separately at the upper portion of the graph, probably due to its high portion of moisture, ashes and total titrateable acidity (TTA) (Table 4.2 and 4.5).

The trait starch was highly associated with the seventh and ninth PCs, which explained a lower portion of genetic variation (5 and 2%, respectively). Shittu *et al.* (2007) observed the

same result. The traits total and free cyanide were highly associated with the tenth and 13th PCs, eigenvectors -0.41 and 0.42, respectively (Table 4.5), showing low importance for the discrimination of genotypes. It should be expected that total cyanide was the most responsible for separation of 'bitter' and 'sweet' cassavas in the graphic. However, this trait was one of the least important to explain genetic variation. A explanation is that the variation of total cyanide among 'bitter' cassava genotypes was wider than within the group of 'sweet' cassava (Table 4.2) and the high coefficient of variation detected for total cyanide measures (93%, Table 4.1).

The cluster analysis represented in Figure 4.2 pointed out that the 13 chemical traits of roots of the 46 genotypes of 'bitter' and 'sweet' cassava allowed classification of the genotypes into two distinct groups that coincide with their classification into 'bitter' or 'sweet' cassava, with the exception of CPATU 034 (S). There was the formation of three subgroups, which varied in number of genotypes from three to 28 (Figure 4.2). Members of group I were characterized by low contents of free and total cyanide, glucose and primarily sucrose when compared with members of groups II and III, that generally have higher contents of free and total cyanide. Moreover, genotypes allocated in groups II and III also presented higher contents of TC, with the exception of CPATU 514, CPATU 190 and CPATU 465. Group II was formed by genotypes CPATU 475, CPATU 465 and CPATU 466, and it was the most divergent group.

The groups of genotypes were determined by Pseudot2 methods and the validity of this hierarchical clustering method was then confirmed by the coefficient of cophenetic correlation of the dendrogram (0.71), which revealed adequate adjustment between graphic representation (Figure 4.2) and the matrix of genetic dissimilarity (Appendix Table 4.2). The result of the hierarchical clustering method corroborated the graphical dispersion of genotypes (Figure 4.1 and 4.2) according to PCs technique and highlighted the potential of chemical traits of cassava root to separate genotypes of 'bitter' and 'sweet' cassava.

Perhaps these results reflected the effects of years of selection that farmers from the Amazon practice on genotypes of cassava, and that include not only selection based on cyanide contents, but also on traits that affect the final product. Clearly, this selection was based on farmers perception, since they do not have means to measure the traits. Considering research programs, due to differences in usage, 'sweet' and 'bitter' cassava genetic breeding programs are specific for each type (Ceballos *et al.* 2004).

REFERENCES

- AOAC (1997) **Official methods of analysis of the Association of the Official Analytical Chemists International**. 16th edn, Patricia Cunniff, Washington, 1683p.
- Burns AE, Gleadow RM, Zacarias AM, Cuambe CE, Miller RE and Cavagnaro TR (2012) Variations in the chemical composition of cassava (*Manihot esculenta* Crantz) leaves and roots as affected by genotypic and environmental condition. **Journal of Agricultural and Food Chemistry** **60**: 4946-4956.
- Carvalho LJCB, Vieira EA, Fialho JF and Souza CRB (2011) A genomic assisted breeding program for cassava to improve nutritional quality and industrial traits of storage root. **Crop Breeding and Applied Biotechnology**. **11**: 289-296.
- Ceballos H, Sánchez T, Chávez AL, Iglesias C, Debouck D, Mafla G and Tohme J (2006) Variation in crude protein content in cassava (*Manihot esculenta* Crantz) roots. **Journal of Food Composition and Analysis**. **19**: 589–593.
- Ceballos H, Iglesias CA, Pérez JC and Dixon AGO (2004) Cassava breeding: opportunities and challenges. **Plant Molecular Biology** **56**: 503-516.
- Charrad M, Ghazzali N, Boiteau V and Niknafs A (2011) **NbClust: An examination of indices for determining the number of clusters. R package version 1.4**. Available at <<https://cran.r-project.org/web/packages/NbClust/index.html>>. Accessed on April 3th, 2017.
- Cruz CD (2006) **Programa Genes: análise multivariada e simulação**. Viçosa: UFV.
- Essers AJA, Bosveld M, Grift RMVD and Voragen AGJ (1993) **Assay for the cyanogens content in cassava products**. Department of food Science, Wageningen. Agricultural University, Netherlands.
- Feltran JC, Lemos LB and Vieites RL (2004) Technological quality and utilization of potato tubers. **Scientia Agricola** **61**: 598-603.
- Food and Agriculture Organization of the United Nations – FAOSTAT. (2014) Production, crops. Available at <<http://faostat3.fao.org/browse/Q/QC/E>> Accessed on April 3th, 2017.
- Grossman J and Freitas AC (1950) Determinação do teor de matéria seca pelo peso específico em raízes de mandioca. **Revista Agronômica** **14**: 75-80.
- Gu B, Yao Q, Li K and Chen S (2013) Change in physicochemical traits associated with genotypes and environmental factors. **Starch**. **65**: 253-263.
- Jolliffe IT (1973) Discarding variables in a principal component analysis. II: real data. **Journal of the Royal Statistical Society. Series C (Applied Statistics)** **22**: 21-31.

- Lebot V (2009). **Tropical Root and Tuber Crops, Cassava, Sweet Potato, Yams and Aroids. Crop Production Science in Horticulture 17**. CABI, Wallingford.
- Léotard G, Duputié A, Kjellberg F, Douzery EJP, Debain C, Granville JJ and McKey D (2009) Phylogeography and the origin of cassava: new insights from northern rim of the Amazonia basin. **Molecular Phylogenetics and Evolution 53**: 329-334.
- McKey D, Cavagnaro TR, Cliff J and Gleadow R (2010) Chemical ecology in coupled human and natural systems: people, manioc, multitrophic interactions and global change. **Chemoecology 20**: 109-133.
- Minitab Inc, **Minitab**, State College, PA. 2008.
- Oliveira EJ, Aud FF, Morales CFG, Oliveira SAS, and Santos VS (2016) Non-hierarchical clustering of *Manihot esculenta* Crantz germplasm based on quantitative traits. **Revista Ciência Agronômica 47**: 548-555.
- Padonou W, Mestres C and Nago MC (2005) The quality of boiled cassava roots: instrumental characterization and relationship with physicochemical properties and sensorial properties. **Food Chemistry. 89**: 261-270.
- Rodriguez-Amaya DB (2001) **A guide to carotenoid analysis in food**. OMNI Research, Washington, USA.
- Sánchez T, Ceballos H, Dufour D, Ortiz D, Morante N, Calle F, Felde T Z, Domínguez M and Davrieux F (2014) Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and Hunter color techniques. **Food Chemistry 151**: 444-451.
- Shittu TA, Sanni LO, Awonorin SO, Maziya-Dixon B and Dixon A (2007) Use of multivariate techniques in studying the flour making properties of some CMD resistant cassava clones. **Food Chemistry 101**:1606-1615.
- Sokal RR and Rohlf FJ (1962) The comparison of dendrograms by objective methods. **Taxon 11**:33-40.
- Stitt M, Lilley RMC, Gerhard R and Heldt HW (1989) Metabolite levels in specific cells and subcellular compartments of plant leaves. **Methods in Enzymology. 174**: 518-552.
- Vieira EA, Fialho JF, Carvalho LJC, Malaquias JV and Fernandes FD (2015) Avaliação de genótipos de mandioca industriais em área de Cerrado do Noroeste de Minas Gerais. **Revista Ceres 62**: 453-459.

Table 4.1. Descriptive analysis of 13 chemical traits of the root of 46 genotypes of cassava (*Manihot esculenta* Crantz) sampled in the Brazilian Amazon and maintained in a germplasm bank from Eastern Amazon, Brazil. Minimum (Min.), maximum (Max.) and mean values, amplitude, standard deviation, coefficient of variation (CV %), and normality test.

Characters	Min.	Max.	Mean values	Amplitude	Standard deviation	CV (%)	Normality test
TTA	1.70	4.37	2.43	2.67	0.55	22	0.95
Glucose	0.04	0.38	0.17	0.35	0.09	52	0.98**
Fructose	0.03	0.25	0.08	0.23	0.05	54	0.91**
Sucrose	0.07	0.77	0.38	0.70	0.22	58	0.97**
TSS	1.94	5.60	3.18	3.65	0.73	23	0.97
TC	1.71	18.41	5.68	16.69	3.88	68	0.91**
Total cyanide	29.80	934.40	228.70	904.60	213.60	93	0.91**
Free cyanide	5.78	183.66	64.68	177.88	53.41	82	0.94**
Ashes	1.35	3.42	2.21	2.06	0.51	23	0.98
pH	5.77	6.78	6.36	1.01	0.21	3	0.96
Crude protein	0.18	2.27	0.88	2.08	0.48	54	0.94**
Moisture	59.18	79.22	67.67	20.04	5.20	8	0.98
Starch	19.79	34.07	27.26	14.28	3.74	14	0.99

TTA: total titrateable acidity (%); glucose (%); fructose (%); sucrose (%); TSS: total soluble solids (%); TC: total carotenoids ($\mu\text{g g}^{-1}$); Total cyanide (mg kg^{-1}); Free cyanide (mg kg^{-1}); ashes (%); pH; Crude protein (%); Moisture (%); Starch (%). **, $P < 0.01$ by Ryan-Joiner test.

Table 4.2. Chemical traits of the root of genotypes of ‘bitter’ and ‘sweet’ cassava (*Manihot esculenta* Crantz) sampled in the Brazilian Amazon and that belong to the germplasm bank in Eastern Amazon, Brazil. Traits were measured on humid basis of roots.

Genotypes	Moisture (%)	Ashes (%)	TSS*	TTA*	pH	Total cyanide (mg kg ⁻¹)	Free cyanide (mg kg ⁻¹)
‘Bitter’ cassava							
CPATU 016	63.83	1.99	3.40	1.83	6.21	202.07	70.86
CPATU 507	65.73	2.04	4.27	1.71	6.58	324.03	88.36
CPATU 348	64.44	2.29	3.18	2.44	6.39	227.29	45.88
CPATU 461	72.19	2.28	3.73	2.40	6.23	238.21	56.28
CPATU 160	60.70	2.33	2.77	3.07	6.49	311.63	73.78
CPATU 140	62.68	1.57	3.28	2.37	6.41	123.78	47.51
CPATU 462	72.64	2.14	3.87	1.72	6.23	259.45	97.35
CPATU 026	64.42	3.42	3.84	3.25	6.13	224.18	80.60
CPATU 466	68.62	2.37	3.67	2.41	5.89	226.03	78.56
CPATU 516	65.20	1.75	3.45	2.08	6.38	89.44	75.23
CPATU 513	67.19	2.08	3.87	2.22	6.36	934.37	150.48
CPATU 475	68.78	2.07	3.80	2.29	5.77	85.70	36.18
CPATU 509	64.62	1.66	3.73	3.22	6.30	222.88	43.80
CPATU 458	68.31	2.27	3.87	2.26	6.43	548.94	91.63
CPATU 463	65.79	1.82	3.01	2.52	6.17	234.58	62.13
CPATU 141	66.49	1.69	2.93	2.69	6.31	207.52	66.68
CPATU 514	60.18	1.54	3.73	2.01	6.54	496.34	37.93
CPATU 190	65.57	1.36	3.67	1.76	6.64	453.35	183.66
CPATU 465	71.18	2.45	3.33	2.08	5.85	288.74	70.97
CPATU 375	69.78	2.82	4.00	1.76	6.39	93.43	36.18
CPATU 515	64.00	2.81	4.13	3.26	6.07	539.79	144.31
CPATU 229	62.89	1.51	3.80	2.62	6.50	315.02	66.58
CPATU 530	60.18	1.47	4.26	2.20	6.36	828.47	182.67
CPATU 193	66.70	3.25	5.60	2.47	6.45	535.72	146.99
CPATU 327	66.69	1.86	3.07	1.82	5.95	431.07	176.87
CPATU 092	64.01	1.74	2.93	2.28	6.33	123.27	50.92
CPATU 404	62.91	1.92	2.73	1.91	6.71	101.13	53.58
CPATU 499	60.07	1.87	2.53	1.84	6.35	179.65	88.40
CPATU 500	68.63	2.44	2.53	2.22	6.23	402.61	146.37
CPATU 219	59.18	1.65	3.40	1.70	6.50	479.22	168.82
‘Sweet’ cassava							
CPATU 179 (S)	62.63	3.11	1.94	2.91	6.42	94.28	77.25
CPATU 280 (S)	74.63	2.84	2.53	1.76	6.78	40.20	15.68
CPATU 241 (S)	72.53	2.82	2.93	2.43	6.49	45.45	14.42
CPATU 271 (S)	74.90	2.14	2.48	2.44	6.47	51.60	12.15
CPATU 019 (S)	73.70	2.59	2.82	2.48	6.69	49.24	10.38
CPATU 232 (S)	69.47	3.15	2.49	2.81	6.41	29.81	9.25
CPATU 309 (S)	75.52	2.36	2.92	2.59	6.44	45.87	9.65
CPATU 183 (S)	72.20	1.94	2.43	2.55	6.38	37.17	9.51
CPATU 070 (S)	70.57	2.07	2.45	2.24	6.60	32.74	7.73
CPATU 281 (S)	75.31	2.71	2.44	2.38	6.44	53.67	9.89
CPATU 283 (S)	79.22	2.97	2.46	4.37	6.21	39.94	13.09
CPATU 274 (S)	72.31	2.04	2.48	2.39	6.44	51.35	5.78
CPATU 115 (S)	77.90	2.62	2.45	3.47	6.60	36.38	14.75
CPATU 195 (S)	61.41	1.98	2.01	2.77	6.40	48.06	7.12
CPATU 136 (S)	74.07	2.06	2.99	3.06	6.47	62.97	14.00
CPATU 034 (S)	62.27	2.01	2.07	2.88	6.42	75.65	25.23

TSS*: total soluble solids; TTA*: total titratable acidity

Table 4.3. Chemical traits of the root of genotypes of ‘bitter’ and ‘sweet’ cassava (*Manihot esculenta* Crantz) sampled in the Brazilian Amazon and that belong to the germplasm bank in Eastern Amazon, Brazil. Traits were measured on humid basis of roots.

Genotypes	TC* ($\mu\text{g g}^{-1}$)	Crude protein (%)	Glucose (%)	Fructose (%)	Sucrose (%)	Starch (%)
Bitter cassava						
CPATU 016	18.41	0.46	0.19	0.07	0.76	25.20
CPATU 507	14.09	0.64	0.24	0.07	0.51	26.32
CPATU 348	11.34	0.89	0.21	0.05	0.77	28.00
CPATU 461	11.24	1.19	0.24	0.07	0.74	24.63
CPATU 160	10.88	0.86	0.18	0.03	0.26	31.40
CPATU 140	9.44	0.80	0.26	0.07	0.30	24.07
CPATU 462	8.48	1.30	0.38	0.10	0.47	31.77
CPATU 026	8.48	0.63	0.24	0.05	0.63	27.74
CPATU 466	7.71	1.28	0.32	0.14	0.62	30.08
CPATU 516	7.39	0.55	0.28	0.18	0.43	31.12
CPATU 513	7.17	0.83	0.30	0.08	0.44	29.43
CPATU 475	6.00	1.10	0.34	0.26	0.53	25.76
CPATU 509	6.00	2.01	0.16	0.06	0.62	33.09
CPATU 458	5.65	2.27	0.22	0.04	0.48	30.00
CPATU 463	5.41	0.63	0.30	0.12	0.41	24.91
CPATU 141	5.01	1.08	0.11	0.07	0.61	26.61
CPATU 514	3.92	1.00	0.22	0.09	0.49	29.56
CPATU 190	3.84	1.94	0.17	0.12	0.45	27.48
CPATU 465	2.97	0.91	0.35	0.20	0.34	28.86
CPATU 375	1.71	1.82	0.23	0.08	0.39	34.07
CPATU 515	n.d.	0.83	0.26	0.10	0.78	26.22
CPATU 229	n.d.	1.10	0.20	0.05	0.38	32.32
CPATU 530	n.d.	1.09	0.19	0.04	0.46	25.76
CPATU 193	n.d.	1.83	0.17	0.05	0.62	28.02
CPATU 327	n.d.	0.70	0.19	0.06	0.64	33.00
CPATU 092	n.d.	1.36	0.12	0.09	0.39	30.00
CPATU 404	n.d.	0.73	0.12	0.06	0.30	27.73
CPATU 499	n.d.	0.55	0.10	0.03	0.29	32.53
CPATU 500	n.d.	1.09	0.19	0.08	0.77	30.84
CPATU 219	n.d.	0.65	0.17	0.04	0.47	32.25
Sweet cassava						
CPATU 179 (S)	4.38	0.46	0.08	0.07	0.14	21.19
CPATU 280 (S)	3.87	0.45	0.12	0.13	0.12	28.81
CPATU 241 (S)	3.42	0.55	0.04	0.06	0.09	22.09
CPATU 271 (S)	3.26	0.82	0.05	0.04	0.11	21.53
CPATU 019 (S)	3.03	0.27	0.04	0.06	0.11	19.79
CPATU 232 (S)	2.95	0.18	0.05	0.06	0.08	20.41
CPATU 309 (S)	2.82	0.46	0.12	0.11	0.16	24.07
CPATU 183 (S)	2.78	0.53	0.05	0.05	0.09	24.34
CPATU 070 (S)	2.63	0.45	0.09	0.09	0.10	26.33
CPATU 281 (S)	2.62	0.44	0.11	0.12	0.17	24.00
CPATU 283 (S)	2.61	0.36	0.12	0.11	0.15	23.79
CPATU 274 (S)	2.51	0.48	0.11	0.06	0.16	27.34
CPATU 115 (S)	2.50	0.53	0.05	0.06	0.11	22.38
CPATU 195 (S)	2.30	1.10	0.09	0.05	0.17	21.82
CPATU 136 (S)	2.08	0.64	0.12	0.10	0.14	28.02
CPATU 034 (S)	n.d.	0.92	0.09	0.11	0.28	29.71

TC*: total carotenoids; n.d.: not determined.

Table 4.4. Estimatives of correlation coefficient of Spearman among 13 chemical traits of the root measured in 46 genotypes of cassava (*Manihot esculenta* Crantz) sampled in the Brazilian Amazon and maintained in a germplasm bank in the Eastern Amazon, Brazil.

Traits	TTA	Glucose	Fructose	Sucrose	TSS	TC	Free cyanide	Ashes	pH	Crude protein	Moisture	Starch
Glucose	-0.33											
Fructose	-0.15	0.40**										
Sucrose	-0.23	0.70**	0.06									
TSS	-0.31	0.75**	0.02	0.70**								
TC	-0.04	0.40**	0.14	0.22	0.24							
Free cyanide	-0.40**	0.57**	-0.09	0.70**	0.63**	0.04						
Ashes	0.40**	-0.11	0.15	-0.20	-0.18	0.06	-0.22					
pH	-0.10	-0.53**	-0.25	-0.52**	-0.26	-0.14	-0.28	-0.09				
Crude protein	-0.14	0.47**	-0.04	0.60**	0.60**	-0.01	0.45**	-0.40**	-0.23			
Moisture	0.13	-0.10	0.35	-0.40**	-0.22	0.05	-0.46**	0.52**	0.05	-0.32		
Starch	-0.40**	0.46**	0.01	0.50**	0.42**	-0.13	0.51**	-0.35	-0.15	0.52**	-0.35	
Total cyanide	-0.33	0.65**	-0.14	0.80**	0.73**	0.07	0.90**	-0.32	-0.23	0.61**	-0.50**	0.52**

TTA: total titrateable acidity (%); glucose (%); fructose (%); sucrose (%); TSS: total soluble solids (%); TC: total carotenoids ($\mu\text{g g}^{-1}$); total cyanide (mg kg^{-1}); free cyanide (mg kg^{-1}); ashes (%); pH; crude protein (%); moisture (%); starch (%). **, $p < 0.01$ correlation;

Table 4.5. Estimatives of eigenvalues and eigenvectors associated with the principal components and 13 chemical traits of the root of cassava (*Manihot esculenta* Crantz) obtained for 46 genotypes sampled in the Brazilian Amazon and maintained in a germplasm bank in Eastern Amazon, Brazil.

Characters	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10	PC 11	PC 12	PC 13
TTA	-0.19	-0.17	-0.52	-0.05	-0.26	0.51	0.24	-0.24	0.4	-0.14	-0.13	0.1	-0.09
Glucose	0.35	-0.36	0.13	0.04	0.02	-0.07	-0.01	-0.22	0.22	0.29	0.3	-0.22	-0.67
Fructose	0.04	-0.51	0.38	-0.31	-0.01	-0.15	0.23	-0.36	-0.06	-0.44	-0.2	-0.01	0.25
Sucrose	0.38	-0.13	-0.15	0.2	0.09	0.14	-0.14	0.32	0.21	-0.25	-0.6	-0.4	-0.16
TSS	0.35	-0.08	-0.22	0.03	0.45	0.07	0.11	-0.29	-0.04	0.48	-0.4	0.32	0.15
TC	0.08	-0.25	-0.13	0.81	0.15	0.13	0.02	0.06	0.17	0.22	0.2	0.21	0.25
Free cyanide	0.36	0.16	-0.23	-0.09	-0.18	-0.37	0.05	0.14	0.03	-0.41	0.08	0.6	-0.23
Ashe	-0.18	-0.3	-0.53	-0.18	0.15	-0.23	-0.57	-0.23	-0.24	-0.16	0.21	-0.11	0.08
pH	-0.19	0.49	0.07	0.15	0.46	-0.1	-0.02	-0.38	0.19	-0.4	0.22	-0.16	-0.28
Crude protein	0.29	0.07	-0.06	-0.24	0.44	0.55	0.16	0.06	-0.36	-0.23	0.39	-0.01	0.01
Moisture	-0.25	-0.31	-0.07	-0.22	0.5	-0.14	0.08	0.6	0.38	-0.04	-0.07	0.01	-0.08
Starch	0.3	0.11	0.22	-0.28	0.02	0.26	-0.64	-0.03	0.48	-0.03	-0.04	0.05	0.21
Total cyanide	0.36	0.18	-0.29	-0.02	-0.02	-0.3	0.3	-0.01	0.32	-0.04	0.23	-0.5	0.42
Eigenvalues	4.8	2.07	1.29	1.13	0.87	0.81	0.55	0.39	0.38	0.24	0.23	0.11	0.09
Accumulated eigenvalues	4.8	6.87	8.16	9.29	10.16	10.97	11.52	11.91	12.25	12.53	12.76	12.87	12.96
Variance	0.37	0.15	0.10	0.09	0.07	0.06	0.05	0.03	0.02	0.02	0.02	0.01	0.01
Accumulated variance	0.37	0.52	0.62	0.71	0.78	0.84	0.89	0.92	0.94	0.96	0.98	0.99	1.0

TTA: total titrateable acidity (%); glucose (%); fructose (%); sucrose (%); TSS: total soluble solids (%); TC: total carotenoids ($\mu\text{g g}^{-1}$); total cyanide (mg kg^{-1}); free cyanide (mg kg^{-1}); ashes (%); pH; crude protein (%); moisture (%); starch (%). Principal components (PCs). Scaling method used: correlation matrix.

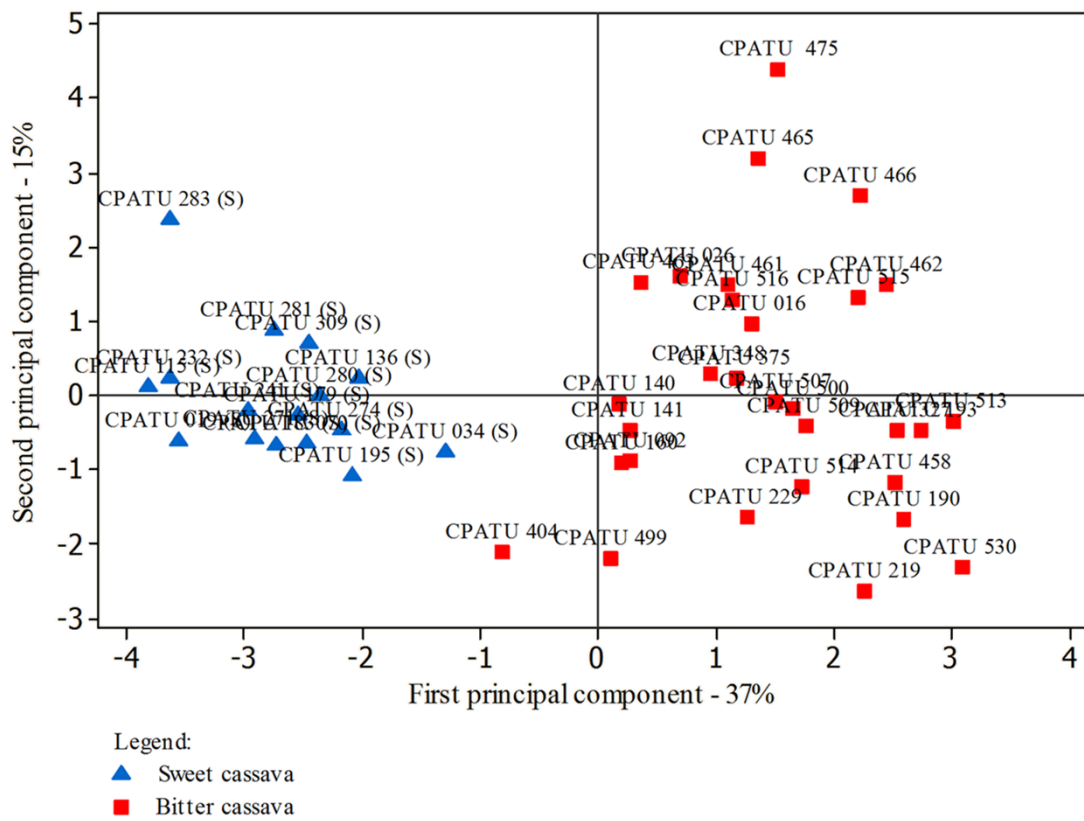


Figure 4.1. Graphical dispersion of 46 genotypes of cassava (*Manihot esculenta* Crantz) sampled in the Amazon and analyzed with 13 chemical traits of the root based on the first two principal components. Scaling method used: correlation matrix.

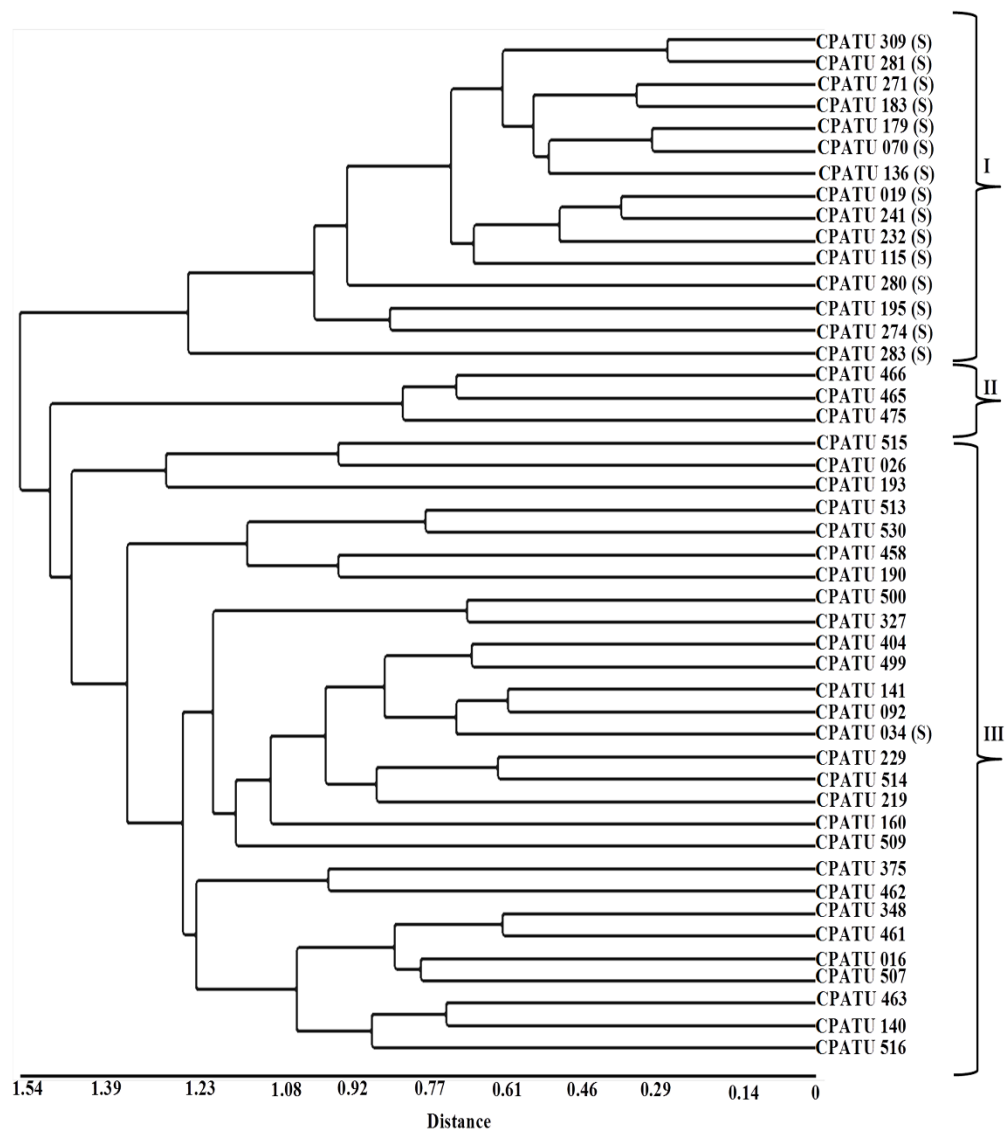


Figure 4.2. Dendrogram generated by UPGMA method based on standardized Euclidean mean distances as measure of dissimilarity, and obtained for 46 genotypes of cassava (*Manihot esculenta* Crantz) sampled in the Amazon and analyzed with 13 chemical traits of the root. Coefficient of cophenetic correlation ($r = 0.71$). Groups were determined based on 'NbClust' package of software R 3.0.3 (R Core Team) using Pseudot2. (S) indicates 'sweet' genotypes of cassava.

CONCLUSÃO GERAL

Os acessos do banco de germoplasma da Amazônia Oriental apresentaram variação genética para descritores quantitativos da mandioca. Foram identificados dois SNPs no gene *psy2* importantes para a síntese de carotenóides, por promovem trocas de aminoácidos na proteína PSY2, o que conseqüentemente indicou que as substituições de aminoácidos afetam o sítio ativo da proteína, isso mostra que a presença do polimorfismo está associada com tonalidade da raiz de mandioca. A partir da aplicação de GWAS foi encontrado um SNP posicionado na região do locus Manes.01G048600, e sua atividade biológica está associada com a enzima glucose-6-phosphate 1-epimerase, que ligada ao metabolismo de carboidrato. Portanto os SNPs podem representar um importante marcador para seleção assistida em programas de melhoramento genético visando o aumento do conteúdo de amido e a biofortificação das raízes.

Este estudo também mostrou acessos com baixo conteúdo de cianeto e níveis moderados de carotenóides nas raízes. Além disso, houve uma clara separação de acessos de mandioca “doce”, “brava” e “açucaradas” com base em técnicas multivariadas usando os caracteres químicos da raiz e a partir dos dados genotípico obtidos da genotipagem por sequenciamento.

APPENDIX

Table 1 Identification of genotypes of ‘bitter’ and ‘sweet’ cassava (*Manihot esculenta* Crantz) sampled in the Brazilian Amazon and that are being maintained in a germplasm bank in the Eastern Amazon.

N ^o *	Genotypes	State	City	Year of sample
‘Bitter’ cassava				
1	CPATU 016	PA	Altamira	1950
2	CPATU 507	PA	Paragominas	2010
3	CPATU 348	PA	Santarém	2001
4	CPATU 461	AP	Pedra Branca do Amaparí	1996
5	CPATU 160	PA	Bragança	1995
6	CPATU 140	AM	Tabatinga	1993
7	CPATU 462	PA	Curuçá	2010
8	CPATU 026	AM	-	1950
9	CPATU 466	PA	Curuçá	2010
10	CPATU 516	PA	Tomé Açu	2010
11	CPATU 513	PA	Capitão Poço	2010
12	CPATU 475	PA	Curuçá	2010
13	CPATU 509	PA	-	2010
14	CPATU 458	AP	Oiapoque	1993
15	CPATU 463	PA	Acará	2010
16	CPATU 141	AM	Tabatinga	1993
17	CPATU 514	PA	Capitão Poço	2010
18	CPATU 190	AP	Ferreira Gomes	1996
19	CPATU 465	PA	Terra Alta	2008
20	CPATU 375	PA	Moju	2002
21	CPATU 515	PA	Altamira	2010
22	CPATU 229	PA	Nova Timboteua	1998
23	CPATU 530	PA	Bragança	1960
24	CPATU 193	PA	Bragança	1996
25	CPATU 327	PA	Santarém	2001
26	CPATU 092	BA	Cruz das Almas	1989
27	CPATU 404	PA	Santa Maria do Pará	2005
28	CPATU 499	PA	Bragança	2010
29	CPATU 500	PA	Patauateua	2010
30	CPATU 219	PA	Castanhal	1998
‘Sweet’ cassava				
31	CPATU 179 (S)	AP	Pedra Branca do Amaparí	1996
32	CPATU 280 (S)	PA	Dom Eliseu	2000
33	CPATU 241 (S)	PA	São Francisco do Pará	2000
34	CPATU 271 (S)	PA	Rondon do Pará	2000
35	CPATU 019 (S)	PA	Ilha do Mosqueiro	1950
36	CPATU 232 (S)	PA	Rondon do Pará	1999
37	CPATU 309 (S)	PA	Santarém	2001
38	CPATU 183 (S)	AP	Pedra Branca do Amaparí	1996
39	CPATU 070 (S)	PA	Santana do Araguaia	1970
40	CPATU 281 (S)	PA	Rondon do pará	2000
41	CPATU 283 (S)	PA	Ulianópolis	2000
42	CPATU 274 (S)	PA	Paragominas	2000
43	CPATU 115 (S)	AP	-	1993
44	CPATU 195 (S)	-	-	1996
45	CPATU 136 (S)	AM	Tabatinga	1993
46	CPATU 034 (S)	PA	Marabá	1996

*The numbers indicate the genotypes in the supplementary table of genetic dissimilarity matrix. Null spaces represent unknown origin of sample. AM: Amazonas, AP: Amapá, BA: Bahia, PA: Pará

Supplementary Table 2 Dissimilarity matrix of analysis of 13 chemical traits of the root of 46 genotypes of cassava (*Manihot esculenta* Crantz) sampled in the Brazilian Amazon and maintained in a germoplasm bank from Eastern Amazon, Brazil.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
1																										
2	.77																									
3	.73	.80																								
4	.87	.88	.61																							
5	1.2	1.1	.84	1.2																						
6	1.0	.84	.81	.93	.92																					
7	1.3	1.0	1.0	.97	1.3	1.1																				
8	1.3	1.3	.91	.99	1.1	1.3	1.3																			
9	1.2	1.3	1.0	.85	1.4	1.2	.73	1.0																		
10	1.2	1.0	1.1	1.1	1.2	.93	.94	1.4	.96																	
11	1.7	1.3	1.4	1.7	1.4	1.4	1.2	1.5	1.3	1.4																
12	1.7	1.7	1.6	1.4	2.0	1.5	1.3	1.7	.90	1.1	1.9															
13	1.5	1.4	1.0	1.1	1.0	1.2	1.2	1.3	1.1	1.3	1.5	1.7														
14	1.8	1.4	1.3	1.2	1.4	1.3	1.1	1.6	1.3	1.5	1.0	1.9	.91													
15	1.1	1.0	.88	.80	1.0	.72	.98	1.0	.79	.80	1.2	1.1	1.3	1.4												
16	1.1	1.1	.71	.81	.95	.79	1.2	1.2	1.1	1.0	1.3	1.6	.85	1.1	.85											
17	1.3	0.9	.95	1.1	1.0	.86	1.1	1.4	1.2	.96	1.0	1.7	1.0	1.0	.88											
18	1.6	1.2	1.4	1.4	1.5	1.3	1.2	1.8	1.5	1.3	1.2	1.8	1.3	.93	1.4	1.2	1.0									
19	1.6	1.6	1.5	1.3	1.7	1.4	1.0	1.4	.70	1.0	1.4	.71	1.6	1.6	.86	1.4	1.4	1.6								
20	1.7	1.3	1.2	1.2	1.4	1.4	.95	1.4	1.4	1.2	1.5	1.6	1.1	1.0	1.2	1.2	1.2	1.4	1.3							
21	1.6	1.5	1.2	1.2	1.5	1.5	1.4	.93	1.0	1.4	1.0	1.6	1.4	1.4	1.0	1.2	1.3	1.6	1.3	1.5						
22	1.5	1.2	1.1	1.2	.99	.99	1.2	1.4	1.3	1.1	1.1	1.80	.87	1.1	.83	.62	1.1	1.5	1.1	1.3						
23	1.7	1.4	1.5	1.5	1.5	1.4	1.5	1.7	1.6	1.6	0.7	2.1	1.5	1.0	1.4	1.3	1.0	1.0	1.8	1.7	1.2	1.1				
24	1.9	1.5	1.5	1.4	1.7	1.8	1.5	1.3	1.5	1.8	1.3	2.1	1.4	1.1	1.6	1.5	1.5	1.4	1.8	1.2	1.1	1.4	1.3			
25	1.5	1.5	1.3	1.3	1.4	1.4	1.0	1.4	1.0	1.3	1.0	1.6	1.4	1.3	1.2	1.1	1.2	1.3	1.3	1.4	1.2	1.1	1.1	1.6		
26	1.5	1.3	1.0	1.2	1.1	.97	1.2	1.4	1.1	1.0	1.4	1.6	.90	1.0	.98	.60	.83	1.1	1.3	.98	1.4	.66	1.3	1.5	1.1	
27	1.5	1.2	1.1	1.4	1.1	.96	1.4	1.6	1.6	1.1	1.5	2.0	1.3	1.3	1.1	.90	.90	1.3	1.7	1.2	1.6	.85	1.4	1.7	1.4	
28	1.5	1.4	1.2	1.5	1.0	1.1	1.5	1.5	1.2	1.4	2.0	1.3	1.4	1.2	.98	1.0	1.4	1.6	1.3	1.6	.86	1.4	1.8	1.0		
29	1.5	1.4	1.0	1.1	1.3	1.3	1.1	1.2	1.0	1.2	1.0	1.7	1.2	1.1	1.0	.90	1.1	1.2	1.3	1.2	.99	1.1	1.3	1.4	.68	
30	1.6	1.2	1.3	1.5	1.2	1.3	1.4	1.6	1.5	1.3	1.0	2.1	1.4	1.2	1.3	1.2	.88	1.0	1.7	1.4	1.4	.83	.85	1.5	.87	
31	1.6	1.6	1.3	1.5	1.1	1.2	1.8	1.3	1.7	1.6	1.8	2.0	1.7	1.8	1.1	1.1	1.6	1.8	1.7	1.7	1.6	1.6	1.8	1.9	1.7	
32	1.7	1.4	1.4	1.5	1.4	1.4	1.5	1.7	1.7	1.2	1.8	1.9	1.8	1.8	1.3	1.4	1.4	1.7	1.6	1.3	1.9	1.5	2.1	2.0	1.8	
33	1.6	1.4	1.3	1.3	1.3	1.2	1.7	1.4	1.7	1.5	1.8	2.0	1.7	1.1	1.2	1.1	1.5	1.8	1.7	1.4	1.7	1.4	1.9	1.8	1.8	
34	1.6	1.5	1.3	1.3	1.3	1.1	1.7	1.6	1.8	1.6	1.9	2.0	1.6	1.6	1.2	1.0	1.5	1.7	1.8	1.5	1.8	1.4	1.9	1.9	1.7	
35	1.7	1.5	1.5	1.5	1.5	1.3	1.9	1.6	1.9	1.6	1.9	2.1	1.9	1.9	1.3	1.3	1.6	1.9	1.9	1.7	1.9	1.6	2.0	2.0	2.0	
36	1.7	1.7	1.5	1.5	1.4	1.4	1.9	1.4	1.8	1.7	2.0	2.1	1.9	2.0	1.3	1.3	1.7	2.1	1.8	1.7	1.8	1.7	2.1	2.0	1.9	
37	1.6	1.4	1.3	1.2	1.3	1.1	1.4	1.4	1.4	1.2	1.7	1.6	1.5	1.6	.98	1.0	1.4	1.7	1.4	1.4	1.6	1.3	1.9	1.9	1.7	
38	1.6	1.5	1.3	1.3	1.2	1.0	1.6	1.5	1.6	1.4	1.8	1.9	1.5	1.6	1.1	.95	1.4	1.7	1.7	1.5	1.8	1.3	1.8	2.0	1.6	
39	1.6	1.4	1.3	1.4	1.2	1.0	1.5	1.6	1.6	1.2	1.7	1.8	1.5	1.6	1.1	1.0	1.2	1.6	1.6	1.4	1.8	1.2	1.8	2.0	1.6	
40	1.6	1.5	1.3	1.3	1.4	1.2	1.5	1.4	1.5	1.3	1.8	1.7	1.7	1.7	1.0	1.1	1.5	1.8	1.4	1.4	1.7	1.5	2.0	1.9	1.7	
41	2.1	2.1	1.8	1.7	1.6	1.7	2.0	1.5	1.8	1.8	2.1	2.0	1.9	2.1	1.4	1.5	2.1	2.4	1.8	1.9	1.7	1.9	2.4	2.2	2.2	
42	1.5	1.4	1.2	1.2	1.1	1.0	1.4	1.4	1.5	1.2	1.6	1.8	1.4	1.5	1.0	.90	1.2	1.6	1.5	1.3	1.7	1.1	1.8	1.9	1.5	
43	1.9	1.8	1.5	1.5	1.4	1.4	1.9	1.6	1.9	1.7	2.0	2.2	1.7	1.9	1.4	1.3	1.8	2.0	1.9	1.7	1.8	1.6	2.1	2.1	2.0	
44	1.6	1.6	1.2	1.4	1.1	.98	1.8	1.5	1.6	1.5	1.8	1.9	1.4	1.5	1.1	.92	1.3	1.6	1.8	1.6	1.7	1.2	1.7	2.0	1.7	
45	1.6	1.4	1.2	1.2	1.1	1.1	1.4	1.4	1.4	1.2	1.6	1.7	1.3	1.5	1.0	.94	1.2	1.6	1.4	1.3	1.6	1.1	1.8	1.8	1.6	
46	1.5	1.5	1.1	1.3	.94	1.0	1.5	1.4	1.4	1.1	1.6	1.7	1.1	1.5	1.0	.77	1.1	1.5	1.5	1.3	1.6	1.0	1.7	1.9	1.5	

Supplementary Table2 Continuation...

	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
1																					
2																					
3																					
4																					
5																					
6																					
7																					
8																					
9																					
10																					
11																					
12																					
13																					
14																					
15																					
16																					
17																					
18																					
19																					
20																					
21																					
22																					
23																					
24																					
25																					
26																					
27	.70																				
28	.75	.67																			
29	.96	1.2	1.1																		
30	1.0	1.0	.77	1.0																	
31	1.3	1.1	1.3	1.1	1.6																
32	1.3	1.0	1.4	1.5	1.7	1.3															
33	1.2	1.0	1.3	1.5	1.7	.80	.88														
34	1.2	1.0	1.3	1.5	1.7	1.1	1.0	.47													
35	1.4	1.1	1.5	1.7	1.8	.93	1.0	.38	.54												
36	1.4	1.2	1.4	1.7	1.9	.60	.94	.44	.76	.56											
37	1.1	1.0	1.3	1.4	1.7	1.0	.77	.53	.60	.63	.75										
38	1.0	.91	1.1	1.5	1.6	.97	1.0	.56	.35	.67	.77	.56									
39	.94	.68	1.0	1.4	1.6	1.0	.66	.63	.61	.67	.86	.50	.46								
40	1.2	1.0	1.4	1.4	1.8	.94	.66	.53	.67	.65	.69	.29	.66	.55							
41	1.7	1.8	1.9	1.8	2.3	1.2	1.5	1.2	1.2	1.3	1.0	1.0	1.2	1.3	1.0						
42	.88	.80	.96	1.3	1.4	1.0	.84	.65	.55	.77	1.7	.50	.33	.32	.59	1.2					
43	1.4	1.3	1.6	1.7	2.2	1.0	1.1	.65	.66	.63	.72	.68	.74	.87	.75	.81	.84				
44	.91	.88	1.2	1.5	1.6	.83	1.3	.93	.80	1.0	.97	1.0	.73	.85	1.0	1.5	.84	1.1			
45	.92	.98	1.2	1.3	1.6	1.1	.93	.77	.75	.89	.98	.44	.60	.56	.64	1.0	.48	.73	1.0		
46	.63	.80	.91	1.2	1.3	1.0	1.1	1.0	1.0	1.2	1.1	.97	.86	.76	.99	1.4	.77	1.2	.72	.78	