

**UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS, AMBIENTAIS E BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS AGRÁRIAS
CURSO DE DOUTORADO EM CIÊNCIAS AGRÁRIAS**

**SELEÇÃO PRECOCE E MAPEAMENTO ASSOCIATIVO
PARA AMIDO WAXY EM MANDIOCA**

CÁTIA DIAS DO CARMO

**CRUZ DAS ALMAS / BAHIA
AGOSTO - 2018**

SELEÇÃO PRECOCE E MAPEAMENTO ASSOCIATIVO PARA AMIDO WAXY EM MANDIOCA

Cátia Dias do Carmo

Engenheira Florestal

Universidade Estadual do Sudoeste da Bahia, 2010.

Tese apresentada ao Colegiado do Programa de Pós-Graduação em Ciências Agrárias da Universidade Federal do Recôncavo da Bahia, como requisito parcial para a obtenção do Título de Doutor em Ciências Agrárias (Área de Concentração: Fitotecnia).

Orientador: Dr. Eder Jorge de Oliveira

**CRUZ DAS ALMAS / BAHIA
AGOSTO - 2018**

FICHA CATALOGRÁFICA

C287s	<p>Carmo, Cátia Dias do. Seleção precoce e mapeamento associativo para amido waxy em mandioca / Cátia Dias do Carmo._ Cruz das Almas, BA, 2018. 112f.; il.</p> <p>Orientador: Eder Jorge de Oliveira.</p> <p>Tese (Doutorado) – Universidade Federal do Recôncavo da Bahia, Centro de Ciências Agrárias Ambientais e Biológicas.</p> <p>1.Mandioca – Melhoramento genético. 2.Mandioca – Amido – Análise. I.Universidade Federal do Recôncavo da Bahia, Centro de Ciências Agrárias, Ambientais e Biológicas. II.Título.</p> <p>CDD: 633.682</p>
-------	--

Ficha elaborada pela Biblioteca Universitária de Cruz das Almas – UFRB.
Responsável pela Elaboração – Antonio Marcos Sarmiento das Chagas (Bibliotecário – CRB5 / 1615).
Os dados para catalogação foram enviados pela usuária via formulário eletrônico.

**UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS, AMBIENTAIS E BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS AGRÁRIAS
CURSO DE DOUTORADO**

**SELEÇÃO PRECOCE E MAPEAMENTO ASSOCIATIVO PARA
AMIDO WAXY EM MANDIOCA**

**COMISSÃO EXAMINADORA DA DEFESA DE TESE DE
CÁTIA DIAS DO CARMO**

Realizada em 10 de Agosto de 2018

Dr. Eder Jorge de Oliveira
Embrapa Mandioca e Fruticultura - CNPMF
Examinador Interno (Orientador)

Profa. Dra. Simone Alves Silva
Universidade Federal do Recôncavo da Bahia - UFRB
Examinador Interno

Dra. Massaine Bandeira e Sousa
Embrapa Mandioca e Fruticultura - CNPMF
Examinador Externo

Dr. Alfredo Augusto Cunha Alves
Embrapa Mandioca e Fruticultura - CNPMF
Examinador Externo

Dr. Onildo Nunes de Jesus
Embrapa Mandioca e Fruticultura - CNPMF
Examinador Externo

A todos os que acreditam na pesquisa,
a minha razão Dona Dinha e Seu Antônio (em memória)
e a todos que cultivam o amor **dedico**.

AGRADECIMENTOS

A Universidade Federal do Recôncavo da Bahia - UFRB e aos professores da pós-graduação pela oportunidade da realização do curso de Doutorado. A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pela concessão de bolsa de estudo. A Empresa Brasileira de Pesquisa Agropecuária – Embrapa (Mandioca e Fruticultura) pelo apoio financeiro e infraestrutura. A Fundação de Amparo à Pesquisa do Estado da Bahia – FAPESB pelo apoio financeiro.

Ao meu orientador Eder Jorge pela confiança, serenidade e principalmente paciência. Por sempre primar a construção do conhecimento, o trabalho em equipe, pela exigência e pela disponibilidade.

Aos amigos do Laboratório de Biologia Molecular: em especial a analista Andresa Priscila de Souza Ramos pelas pesquisas, pelas madrugadas, milhares de *emails* e infinitas cotações; aos técnicos Vandeson Rodrigues de Souza e Raimundo Pereira da Silva; aos amigos de outros projetos lane dos Santos Queiroz, Dalma Brito Santos, Paulo Henrique da Silva, Ariana da Silva dos Santos, Julianna Matos da Silva, Eliana Maria Rocha Sousa, Marciene Amorim Rodrigues, Zanon Santana Gonçalves, Zalmar Santana Gonçalves pelo apoio principalmente na extração de DNA das amostras, pela partilha de dúvidas, ensinamentos e amizade.

Aos amigos do Laboratório de Ciência e Tecnologia de Alimentos na pessoa de Jaciene Lopes de Jesus pelo cuidado, atenção, apoio e ensinamentos para avaliação do amido.

A equipe Mandioca com quem tive a sorte e o prazer de participar, crescer e auxiliar no crescimento e formação de tantos outros. Não tenho palavras que possam expressar toda minha gratidão. As ICs companheiras de projeto Luziane Brandão Alves, Priscila Patrícia dos Santos Silva, Luana Ferreira dos Santos, Jocilene dos Santos Pereira e Bruna Leite Vieira do Nascimento, à Ana Claudia Oliveira Barbosa e Danilo Almeida Brito; aos atualmente mestres Letícia Maróstica de Vasconcelos, Paulo Henrique Gomes Alves de Oliveira, Hilçana Gonçalves de Albuquerque, Reizaluamar de Jesus e Emile Lemos Freitas; aos mestrandos Gilvanara Damasceno de Souza e Cristiano Silva dos Santos; aos companheiros de doutorado Leonardo Silva Souza, Camila Santiago Hohenfeld, Gilmara Alvarenga Fachardo Oliveira e Luciano Rogério Braatz de Andrade; aos pós

doutores Ana Carla Brito, Massaine Bandeira e Sousa, Maria Selma Alves Silva Diamantino, Rosa Karla Nogueira Pestana, Rafael Parreira Diniz, Diego Fernando Marmolejo Cortes; a analista Cinara Fernanda Garcia Morales; a IC Jr. Maria de Fátima Fonseca dos Santos; a Alison Borges Vitor; e a toda equipe de campo. Agradeço pelo apoio na obtenção dos dados e por travarmos juntos muitas “brigas” com o ambiente de programação R, Tassel, MrBayes, ImageJ...

A Massaine Bandeira e Sousa (Chefa 2) pessoa fundamental na conclusão deste trabalho. Obrigada pela dedicação, cuidado e pelos valiosos ensinamentos.

A Diogo Pereira Silva de Novais, meu companheiro, meu porto seguro, pela atenção dada a minhas dúvidas, por nunca medir esforços ou distância, pelos momentos de reflexão, leituras, pesquisas e discussões fundamentais no aprimoramento das minhas ideias, pelos conselhos e amparo constante.

A todos os professores que marcaram a minha vida acadêmica em especial a Alessandro de Paula, Avaldo de Oliveira Soares Filho, Christian Dias Cabacinha, Gilmar Correia Silva, Quelmo Silva de Novaes e Edna Lôbo Machado.

A minha bolha do amor; divas do meu coração Claudinha, Dalma, Andresa (Chefa do amor) e Paulo e a minha companheira de madrugada e lágrimas Elaine Silva da Cruz pelos momentos de reflexão, leituras, pesquisas, pelas horas dedicadas ao meu trabalho, pelo amor e a todas descobertas que fizemos juntos. As amigas com quem tive a felicidade/sorte/prazer de encontrar: Letícia Almeida Motta de Moura, Patrícia Reis de Oliveira Silva, Josivania Silveira da Silva, Tatiane Dantas; e aos amigos que peguei emprestado Idália Souza dos Santos, Mateus de Carvalho Silva, Andressa Rodrigues de Oliveira Sousa e Lorena Brito Pimentel Rodrigues dos Santos pela força e apoio durante os anos de doutorado.

A minha família pela compreensão da ausência. Ao meu pai Antônio Rodrigues do Carmo (*em memória*) e a minha mãe Dercely Oliveira Dias do Carmo pelo apoio de sempre. A minha irmã Cassia Dias do Carmo e seu digníssimo Michael Daian Pacheco Ramos, que sempre me abrigaram nos fins de semana e pelas deliciosas discussões sobre ciência geral.

Aos membros da banca, Simone Alves Silva, Massaine Bandeira e Sousa, Alfredo Augusto Cunha Alves e Onildo Nunes de Jesus pelas sugestões para a melhoria desse trabalho.

A todos que a falta de memória não permitiu nomear e que contribuíram de forma direta ou indireta para a realização desse trabalho. **Muito obrigada!!!**

“(...) está cada vez mais claro que um ser vivo é muito mais que a simples soma ou junção das suas partes.”

Diego Bonato

Helder Takashi Imoto Nakaya

Capítulo 13 – Genômica e Biologia de Sistemas

Livro: Ciências genômicas: fundamentos e aplicações. Organizador:
Leandro Marcio Moreira, Sociedade Brasileira de Genética, 2015.

SUMÁRIO

Página

RESUMO

ABSTRACT

INTRODUÇÃO GERAL1

REFERENCIAL TEÓRICO3

ARTIGO 1

ASSOCIAÇÃO DE MUTAÇÕES NO GENE GBSSI COM O FENÓTIPO WAXY EM MANDIOCA.....20

ARTIGO 2

ASSOCIAÇÃO GENÔMICA AMPLA PARA IDENTIFICAÇÃO DE REGIÕES GENÔMICAS ASSOCIADAS AO AMIDO WAXY EM MANDIOCA.....47

ARTIGO 3

IDENTIFICAÇÃO DE GENÓTIPOS WAXY DE MANDIOCA COM USO DO INFRAVERMELHO PRÓXIMO COM TRANSFORMADA DE FOURIER (FT-NIR)81

CONSIDERAÇÕES FINAIS 111

SELEÇÃO PRECOCE E MAPEAMENTO ASSOCIATIVO PARA AMIDO WAXY EM MANDIOCA

Autora: Cátia Dias do Carmo
Orientador: Dr. Eder Jorge de Oliveira

RESUMO: Dada a importância da mandioca (*Manihot esculenta* Crantz) para uso industrial e o recente interesse pelo amido *waxy* e as dificuldades inerentes ao melhoramento desta característica, o objetivo deste trabalho foi identificar estratégias para seleção precoce de genótipos *waxy*. Neste estudo, foram utilizadas três estratégias: 1) avaliação de mutações no gene *Granule-bound starch synthase I* (*GBSSI*) responsável pelo fenótipo *waxy*. Foi realizado um *screening* no germoplasma de mandioca com iniciadores específicos, seguido do sequenciamento total do gene *GBSSI* e avaliação dos pontos de mutação encontrados no sequenciamento individual e conjuntamente por análise discriminante de componentes principais (ADCP). A deleção MeWxEx6-del-C foi identificada como marcador funcional para o fenótipo *waxy* e não foram encontrados genótipos heterozigóticos no germoplasma de mandioca para esta deleção. 2) uso de um painel de marcadores com 20k de *Single Nucleotide Polymorphism* (SNP) para análise de Associação Genômica Ampla (GWAS) para o fenótipo *waxy*. Dez variantes polimórficas foram associadas ao fenótipo localizadas no cromossomo 2 e co-localizados em regiões gênicas que podem direcionar novos estudos e revelar processos biológicos concomitantes. Também foram encontrados efeitos aditivos e epistáticos entre os SNPs significativos; 3) análise de infravermelho próximo com transformada de Fourier (FT-NIR) com o objetivo de associar espectros com componentes da folha presentes em genótipos *waxy* e não-*waxy* que possam distingui-los. Foram avaliadas cinco técnicas de reconhecimento de padrões supervisionado com espectros, sendo que os algoritmos de *Support vector machines with linear kernel* (SVM) e *Bayesian generalized linear model* (BGLM) apresentaram as maiores acurácias de classificação, possibilitando a implementação da seleção precoce de indivíduos em populações segregantes, por meio da seleção indireta por infravermelho próximo.

Palavras-chave: *Manihot esculenta* Crantz, amido ceroso, *GBSSI*, GWAS, infravermelho próximo

EARLY SELECTION AND GENOME-WIDE ASSOCIATION STUDY FOR WAXY STARCH IN CASSAVA

Author: Cátia Dias do Carmo
Advisor: Dr. Eder Jorge de Oliveira

ABSTRACT: Given the importance of cassava (*Manihot esculenta* Crantz) for industrial purposes and the recent interest in waxy starch and the difficulties to improve this characteristic, the goal of this study was to identify strategies for early selection of waxy genotypes. In this study, three strategies were used: 1) Evaluation of mutations in the Granule-bound starch synthase I gene (GBSSI), responsible for the waxy phenotype in several amylaceous crops. A screening was performed in cassava germplasm with specific primers, followed by complete GBSSI gene sequencing and evaluation of the mutation points found in individual sequencing and by Discriminant Analysis Principal Component (DAPC). The deletion MeWxEx6-del-C was identified as a functional marker for the waxy phenotype and heterozygous genotypes were not found in the cassava germplasm for this deletion. 2) use of a panel of markers with 20k Single Nucleotide Polymorphism (SNP) for Genome-wide association studies (GWAS) to the waxy phenotype. Ten polymorphic variants were associated with the phenotype located on chromosome 2 and co-located in gene regions that can drive further studies and reveal concomitant biological processes. Potential additive and epistatic effects were also found among significant SNPs; 3) analysis of Fourier Transform Near-infrared (FT-NIR) in order to associate the spectra with leaf components present in waxy and non-waxy genotypes that can be used to distinguish them. Five techniques of pattern recognition supervised with spectra were evaluated. The algorithms *Support vector machines with linear kernel* (SVM) and *Bayesian generalized linear model* (BGLM) showed the highest classification accuracies, making possible the implementation of the early selection of individuals in segregating populations using indirect selection by near-infrared.

Keywords: *Manihot esculenta* Crantz, waxy starch, GBSSI, GWAS, near-infrared

INTRODUÇÃO GERAL

Dentre as atividades do melhoramento de plantas, a seleção de genótipos e/ou o correto direcionamento dos cruzamentos no processo de introgressão dos alelos de interesse, são pontos chaves para o sucesso dos programas de melhoramento. Em espécies de ciclo longo, como a mandioca, a seleção precisa de características alvo é ainda mais relevante. Assim, determinar metodologias para seleção precoce de caracteres desejáveis é um dos grandes desafios para os melhoristas. No caso da mandioca um ciclo típico de seleção requer um ano para produzir as sementes botânicas provenientes de cruzamentos controlados e de cinco a seis anos consecutivos para avaliação de campo, além de mais três a quatro anos para multiplicação, até o lançamento de uma cultivar (BARANDICA et al., 2016). Com a seleção precoce, espera-se diminuir esse tempo.

Muitas ferramentas auxiliam o melhorista na difícil tarefa de selecionar genótipos desejáveis, principalmente as de difícil identificação e alto custo de avaliação. Dentre elas, podemos citar a seleção assistida por marcadores moleculares de DNA (SAM) e a espectrometria de infravermelho. Análises com marcadores moleculares de DNA podem ser realizadas em qualquer tecido e/ou estágio de desenvolvimento da planta com neutralidade fenotípica e, portanto, ideal para características com alta interação genótipo-ambiente. No entanto, para a seleção de características específicas, é preciso haver conhecimento prévio dos marcadores moleculares de DNA funcionais associados à região causal, e um extenso trabalho para localização e validação desta variante para seu uso eficaz. Ademais, entender a herança da característica de interesse, identificar os genes envolvidos e os efeitos das variações ambientais e genéticas sob o fenótipo geralmente é uma atividade bastante complexa.

Em relação a técnica de espectroscopia de infravermelho próximo, o propósito é uma fenotipagem rápida e precoce, contrapondo uma avaliação mais laboriosa e/ou de maior custo. A técnica é baseada nas vibrações dos átomos da molécula obtidas pela passagem da radiação infravermelha através de uma amostra e pela determinação de qual fração da radiação incidente é absorvida em uma determinada energia (NG; SIMMONS, 2015). No entanto, para seu uso eficaz, as diferenças que se pretendem observar devem ser detectáveis no tecido vegetal avaliado, possibilitando uma distinção com acurácia satisfatória.

No presente trabalho foram utilizadas análises de variantes do DNA genômico e análise por espectrometria de infravermelho de folhas, com foco no amido de mandioca, especificamente com baixo teor de amilose (amido ceroso - *waxy*). O amido *waxy* de mandioca possui propriedades de pasta superiores aos amidos de outras fontes apresentando menor sinérese quando comparado ao amido ceroso de outras culturas e sabor neutro que lhe confere vantagem em relação aos amidos de cereais para a indústria (BREUNINGER; PIYACHOMKWAN; SRIROTH, 2009; SÁNCHEZ et al., 2010). No entanto, o desenvolvimento de novas variedades *waxy* ainda é limitado devido à dificuldade de florescimento dos genótipos elite, presença de alta heterozigose inerentes da cultura (CEBALLOS; KULAKOW; HERSHEY, 2012) e o caráter recessivo do fenótipo (CEBALLOS et al., 2007). Portanto, a identificação da expressão do gene *waxy*, antes do seu armazenamento nas raízes da planta certamente será um avanço importante no estabelecimento de métodos acurados e precoces de predição. Além disso, devido à duração de cada ciclo de seleção recorrente e às limitações dos materiais de plantio nos estágios iniciais, é importante maximizar as possibilidades de selecionar os melhores genótipos no início do processo (BARANDICA et al., 2016). Desta forma o objetivo geral desta tese foi explorar diferentes estratégias para identificação precoce de genótipos com amido ceroso na cultura da mandioca seguindo três estratégias em painéis de genótipos com amido *waxy* e não-*waxy*:

- ✓ *Artigo 1* – análise por marcadores moleculares de DNA referentes ao gene *GBSSI* (*Granule-bound starch synthase I*) e sequenciamento completo do mesmo gene objetivando identificar variantes causais;
- ✓ *Artigo 2* – análise com marcadores moleculares de DNA com cobertura em todo genoma da mandioca por Associação Genômica Ampla (GWAS) objetivando identificar novas variantes;
- ✓ *Artigo 3* – análise por Infravermelho próximo com Transformada de Fourier (FT-NIR) a partir de folhas objetivando identificar modelo precisos de classificação dos fenótipos.

REFERENCIAL TEÓRICO

Características gerais da cultura

Características gerais da espécie

Da família das Euphorbiaceae Juss a espécie cultivada *Manihot esculenta* Crantz é uma planta arbustiva com morfologia e usos bastante diversificados (Figura 1). Sua estrutura morfológica pode variar de genótipo para genótipo em vários aspectos como na folha (número de lóbulos, coloração), caule (cor externa do caule, do córtex, da epiderme e dos ramos terminais), na raiz (cor de polpa, cor da casca, forma) além de diferenças no hábito de crescimento (ereta ou ramificada) e retenção foliar (FUKUDA; GUEVARA, 1998).



Figura 1: Exemplos de variedades de mandioca lançadas pela Empresa Brasileira de Pesquisa Agropecuária - Embrapa. No painel superior três variedades utilizadas para uso industrial e no painel inferior variedades para mesa com demonstração das diferenças morfológicas e de uso entre os genótipos.

Em relação as características fotossintéticas a mandioca é considerada intermediária do tipo C3 e C4 o que lhe confere maior fixação de CO₂ por unidade de água perdida e, portanto, mais adaptada a condições de déficit hídrico. Apesar de não possuir uma anatomia foliar típica das plantas C4, a mandioca possui um complexo enzimático relacionado. Além disso, possui controle estomático eficiente e diminui a

retenção foliar em condições de escassez de água (EL-SHARKAWY, 2007; EL-SHARKAWY; COCK; PORTO, 1989).

A mandioca é uma diploide funcional $2n=36$ ($n=18$) com pequenos e similares cromossomos (DE CARVALHO; GUERRA, 2002), e no recente sequenciamento da mandioca ficou claro o fenômeno de paleotetraploidia na cultura. Observando a distribuição genômica de parálogos, os grupos de ligação da mandioca compreendem cinco pares de cromossomos homólogos e dois grupos de quatro cromossomos que sofreram uma série de quebras e fusões envolvendo homólogos (BREDESON et al., 2016).

Adicionalmente, as raízes da mandioca podem ter conteúdo de ácido cianídrico variando de 5 a mais de 800 ppm mg Kg^{-1} de raiz fresca determinando sua utilização na alimentação (mandiocas mansas – até 50 ppm mg Kg^{-1}) ou na produção de fécula e farinha (mandioca brava acima de 50 ppm mg Kg^{-1}). No contexto do melhoramento da cultura, variedades para a indústria, devem possuir características como produtividade, alto teor de matéria seca e amido, enquanto para consumo humano as características de cozimento, sabor e coloração da raiz em algumas regiões são mais importantes (CEBALLOS; KULAKOW; HERSHEY, 2012) (Figura 1).

A mandioca é alógama (polinização cruzada) com flores masculinas e femininas presentes na mesma planta (monoica). As flores femininas se abrem de uma a duas semanas antes das flores masculinas (protoginia) e a floração dos genótipos pode ocorrer entre 4-5 meses após o plantio ou ainda de 8-10 meses após o plantio, dependendo do genótipo (JENNINGS, D. L.; IGLESIAS, 2002) (Figura 2). Além disso, muitos genótipos têm dificuldades de florescimento, em especial os de caule ereto sem ramificação, tornando esta característica um importante entrave na introgressão de alelos (CEBALLOS et al., 2017).



Figura 2: Exemplos de inflorescência na cultura da mandioca. A esquerda inflorescência femininas e masculinas a direita.

A propagação comercial, no entanto, é realizada vegetativamente por secções do caule denominadas popularmente como manivas. Qualquer parte do caule pode ser usada, preferencialmente, as hastes maduras, quando em um corte transversal o diâmetro da medula é igual ou inferior a 50% do diâmetro do caule (CEBALLOS; CALLE, 2007).

Outro entrave importante à expansão e melhoramento da cultura é a necessidade do rápido processamento das raízes, devido a deterioração fisiológica pós colheita (DFPC) que as tornam impróprias para consumo e reduz seu valor comercial dentro de 24-72h (VENTURINI et al., 2016). Além das implicações comerciais, esta característica dificulta as avaliações pós colheita uma vez a quantidade de genótipos avaliadas na colheita é proporcional a disponibilidade de recursos (humanos e materiais) e a complexidade da análise.

Importância econômica e social

A mandioca teve sua origem de domesticação no sudoeste da bacia amazônica (LÉOTARD et al., 2009; OLSEN; SCHAAL, 1999) sendo o Brasil considerando o centro de diversidade da espécie. Esta diversidade de características morfofisiológicas (Figura 1) torna a mandioca uma cultura que se adapta a diferentes condições edafoclimáticas e a diferentes usos. Desta forma, é uma cultura amplamente cultivada nos trópicos (Figura 3) usada como fonte de energia na alimentação humana e animal e para diferentes fins comerciais.

A mandioca tem importante papel na segurança alimentar, principalmente em países do continente africano, devido ao seu baixo custo de implementação como cultura de subsistência e com alta tolerância a seca. No entanto, apesar da sua contribuição na segurança alimentar, estudos na Tanzânia, por exemplo, revelam que o cultivo da mandioca é limitado pela incidências de pragas, falta de mercado, mal processamento das raízes e falta de conhecimento (REINCKE et al., 2018) e portanto ações de melhoramento e difusão de tecnologia são urgentes.

O consumo das raízes frescas é bem diverso entre os países produtores sendo consumida cozida, frita e utilizada numa série de pratos regionais como sopas, caldos e bebidas (BALAGOPALAN, 2002). A produção de alimentos da mandioca em muitos países africanos é trabalhosa, demorada e não industrializada (RAHEEM; CHUKWUMA, 2001). Após processamento das raízes é possível a obtenção de produtos como a farinha, tapioca, polvilho azedo e polvilho doce usados nos preparos dos mais diversos alimentos como pão de queijo e bolos (CHUZEL, 2001). Outras

partes da planta também podem ser utilizadas como alimento. Após desintoxicação, as folhas de mandioca podem fornecer uma importante fonte de alimento de forma segura e nutritiva rica em minerais, proteínas e vitaminas podendo ser usada na alimentação humana e animal (CEBALLOS; KULAKOW; HERSHEY, 2012; LATIF; MÜLLER, 2015).

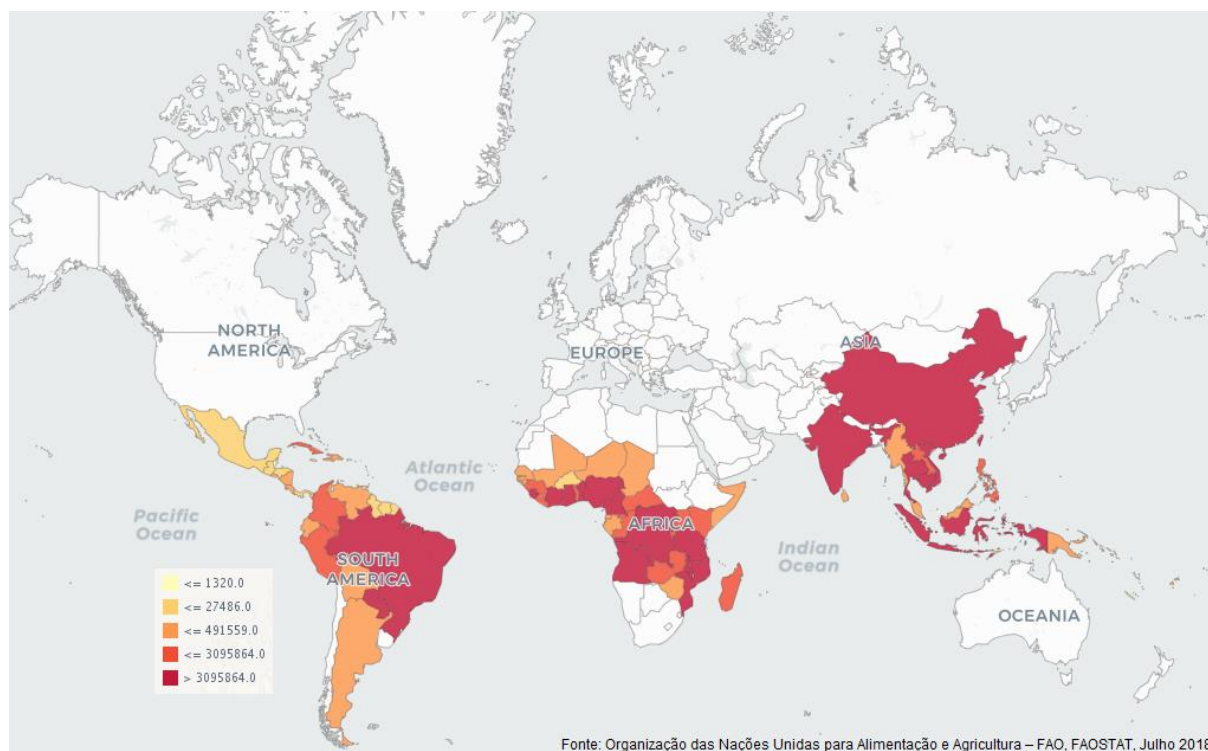


Figura 3: Mapa mundial com destaque para os principais países produtores de mandioca. Diferentes cores indicam produção total no ano de 2016 em toneladas conforme legenda (Fonte: Fundação das Nações Unidas para Alimentação e Agricultura – FAO, 2018).

A produção mundial de 2016 foi de mais de 281 milhões de toneladas com destaque para Nigéria (>57 milhões), Tailândia (>31 milhões), Brasil (>21 milhões) e Indonésia (>20 milhões) (FAOSTAT, 2018). No Brasil a mandioca é produzida de Norte a Sul do país e a estimativa de produção em 2018 deve alcançar cerca de 21 milhões de toneladas (IBGE, 2018). Sua exploração comercial se concentra na venda de raízes para alimentação humana e animal e na produção de amido e farinha (CEBALLOS et al., 2004). Apesar de em algumas localidades ter reputação de “*poor man’s crop*” (REINCKE et al., 2018) no Sul do Brasil a mandioca é cultivada em larga escala com plantio e colheita semi-mecanizados (CHUZEL, 2001).

O amido de mandioca é utilizado principalmente na indústria alimentícia como estabilizador, devido ao sabor neutro e em diferentes aplicações nas indústrias têxtil

e de papel (BREUNINGER; PIYACHOMKWAN; SRIROTH, 2009), representando 22% do consumo de amido do Brasil (ano 2000) (CHUZEL, 2001).

Características gerais do amido

Tipos, propriedades e função do amido

O amido é um polímero insolúvel produzido por plantas e algas basicamente constituído por dois monômeros de glicose: amilose (essencialmente α 1,4-poliglucanos - lineares) e amilopectina (α 1,4-poliglucanos e α -1,6-poliglucanos - ramificados). Nas plantas superiores pode ser sintetizado nos cloroplastos atuando como reserva de curto prazo, usada na fase bioquímica da fotossíntese ou nos amiloplastos como reserva de energia de longo prazo (TAIZ; ZEIGER, 2017). Este último é explorado comercialmente pelo cultivo de sementes (milho, trigo, arroz, sorgo entre outras), tubérculos (batata, batata doce e inhame) e raízes tuberosas (mandioca).

Na indústria, pode ser utilizado como espessante, agente gelificante, agente de volume, agente de retenção de água e adesivo (SINGH et al., 2003). A função do amido na indústria está diretamente relacionada com a estrutura, tamanho dos grânulos do amido e as proporções da amilose e da amilopectina que variam entre espécies ou entre genótipos (SINGH et al., 2003).

Na indústria, modificações físicas e químicas são por vezes necessárias para superar características indesejáveis e/ou tornar o amido adequado para utilizações específicas. A maioria dos amidos "*in natura*" são limitados nas suas aplicações, devido à sua instabilidade sob várias temperaturas e condições de pH com maior parte dos grânulos inertes, insolúveis à temperatura ambiente, altamente resistentes à hidrólise enzimática e, portanto, sem propriedades funcionais específicas (ASHOGBON; AKINTAYO, 2014). Adicionalmente, existe uma resistência ao uso de amidos modificados quimicamente em aplicações alimentares, uma vez que muitos produtos químicos são utilizados neste processo (MOORTHY, 2002). Desta forma, amidos com características diversificadas são objeto do melhoramento das culturas amiláceas, visando diminuir a necessidade de modificações do amido para fins comerciais, além da redução de custos e de resíduo.

Em mandioca, variações em relação ao tipo de grânulos (redondo, truncado cilíndrico, oval, esférico e composto) variando de 4-43 μ m foram reportadas (MOORTHY, 2002). Em relação as proporções de amilose/amilopectina, o teor médio de amilose foi de 20,7% em uma avaliação de 3.272 variedades locais (incluindo 12

parentes selvagens) e 772 clones melhorados de mandioca sem diferença significativa entre variedades locais e clones melhorados (SANCHEZ et al., 2009). No entanto, foram reportadas como resultado de uma série de autofecundações um genótipo com baixo conteúdo de amilose (CEBALLOS et al., 2007) e como resultado de indução por raios gama um genótipo com alto conteúdo de amilose (CEBALLOS et al., 2008), ambos com características de amido diferenciadas em relação aos clones comerciais.

De maneira geral o amido de mandioca possui alta viscosidade, clareza, baixo teor de lipídios e fácil gelatinização (MOORTHY, 2002), tornando-o útil em muitas aplicações alimentares e industriais. Por absorver mais água, o amido de mandioca é usado na indústria frigorífica para amaciar carnes e como agente ligante em salsichas e linguiças. Uma vez que não possui glúten, é utilizado para o preparo de produtos para celíacos ou para padronizar o teor de glúten em diversos alimentos (FOOD INGREDIENTS BRASIL, 2015). O menor teor de lipídios e proteínas confere sabor neutro, com consequentes vantagens em relação aos amidos de cereais para a indústria de alimentos (BREUNINGER; PIYACHOMKWAN; SRIROTH, 2009) tornando o amido de mandioca competitivo em relação a outras culturas.

Rota metabólica do amido

De maneira geral, as enzimas relacionadas à síntese de amido são agrupadas em: ADP-glicose pirofosforilase, sintases, enzimas de ramificação e enzimas desramificadoras do amido. A maioria destas enzimas existem em duas ou mais isoformas com função às vezes indeterminada (PFISTER; ZEEMAN, 2016).

A primeira etapa na biossíntese do amido é a formação de ADP-glicose. A enzima ADP-glicose pirofosforilase catalisa a formação de ADP-glicose a partir de ATP e glicose-1-fosfato com a liberação simultânea de pirofosfato. A próxima etapa é a adição sucessiva de porções glicosil por meio de ligações α -D-1,4 que alongam o polissacarídeo. Amidos sintase transferem a porção glicosil da ADP-glicose para a extremidade não redutora de um primer preexistente de glucano α -D-1,4. As múltiplas isoformas da amido sintase encontrada nos tecidos das plantas são as amidos sintase solúveis (SS), que estão divididas entre as frações granulares e estromais de acordo com a espécie, os tecidos e os estágios de desenvolvimento e as amidos sintase ligadas ao grânulo, localizadas essencialmente no interior da matriz do grânulo (GBSS). Enzimas de ramificação catalisam a formação de pontos de ramificação nas cadeias de glucano por meio de clivagem das ligações α -D-1,4 e a transferência do

oligossacarídeo liberado para um glucano linear, formando uma ligação α -D-1,6 (TAIZ; ZEIGER, 2017) (Figura 4).

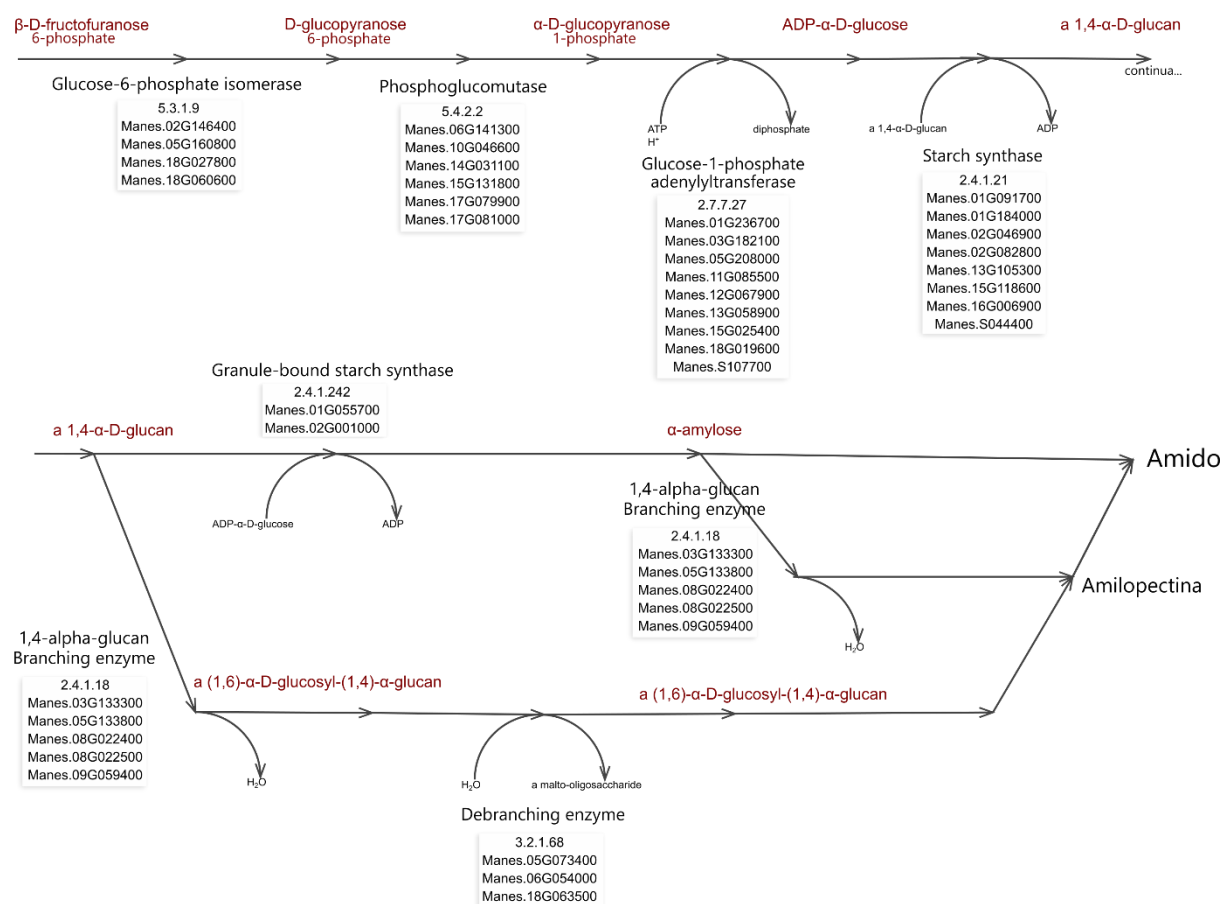


Figura 4: Rota da síntese de amido em mandioca com genes e enzimas correspondentes: alongamento e ramificação do amido.

(Fonte: <https://phytozome.jgi.doe.gov/phytozome> - Identificador: Mesculenta PWY-6724).

Em relação ao amido de armazenamento, a sacarose, sintetizada nas folhas, é transportado para o órgão de armazenamento e importado para o citosol de cada célula onde ocorre sua síntese (KEELING; MYERS, 2010). Em mandioca foram reportadas 270 proteínas relacionadas com a biossíntese do amido, sendo 141 proteínas desempenhando papel no metabolismo de sacarose e amido e as demais 129 proteínas com funções no ciclo de Calvin (SAITHONG et al., 2013). Uma das primeiras vias metabólicas mais completa de mandioca foi reconstruída empregando dados do primeiro sequenciamento do genoma da mandioca (PROCHNIK et al., 2012) conjuntamente com sequências ortólogas de cinco plantas modelo (*Arabidopsis*, arroz, milho, mamona e batata) e complementadas com dados de *microarray* (SAITHONG et al., 2013). No entanto, genes como o GBSSI (AIEMNAKA et al., 2012)

não foram incluídos no trabalho de Saithong et al., (2013). Com o sequenciamento mais completo da cultura, realizado com maior qualidade de montagem (BREDESON et al., 2016) foi disponibilizado no PhytoMine uma rota metabólica de amido (Mesculenta PWY-6724) com as enzimas correspondentes (Figura 4).

Melhoramento genético vegetal assistido por dados genômicos

Técnicas genômicas

O sequenciamento do DNA se desenvolveu rapidamente partindo das primeiras técnicas de sequenciamento concebidas por Maxam e Gilbert (sequenciamento por degradação química - 1977) e por Sanger (sequenciamento método enzimático - 1975) até as plataformas de sequenciamento de larga escala *Next Generation Sequencing* - NGS (2005) e sequenciamento de segunda e terceira geração (2011) onde poucas horas são necessárias para obter um genoma eucarioto completo (OLIVEIRA et al., 2013).

Atualmente, o sequenciamento Sanger utiliza fluoróforos diferentes para cada um dos quatro tipos de didesoxinucleotídeos, que ao serem excitados, emitem luz característica do didesoxinucleotídeo incorporado. Todas as posições deste DNA terão em algum momento, um dos ddNTP incorporado pela DNA polimerase durante a PCR sendo possível a leitura de toda a sequência em tempo reduzido (FIETTO; MACIEL, 2015).

Outro avanço importante em relação as técnicas empregadas na biologia molecular foi a automatização de várias etapas do processo de genotipagem possibilitando a sua utilização em larga escala. Dentre as plataformas podemos citar a genotipagem por sequenciamento (GBS) que, de maneira simplificada, reduz a complexidade do genoma por meio de enzimas de restrição gerando uma biblioteca de fragmentos, onde que, para cada indivíduo são incluídos adaptadores (*barcodes*), permitindo o sequenciamento de centenas de amostras simultaneamente, em um processo chamado “multiplex” (ELSHIRE et al., 2011). Desta forma, milhares de SNPs são identificados, podendo ser utilizados em análises de diversidade genética, identificação de duplicatas, seleção genômica e associação genômica, dentre outras (OLIVEIRA et al., 2013). De posse da matrix de SNPs e de dados de fenotipagem é possível analisar uma associação entre os marcadores e características agrônômicas importantes.

Quando o SNP ou InDel alvo já foi identificado, é possível utilizar técnicas *uniplex* como o *Kompetitive Allele Specific PCR* - KASP para análise de grande

número de indivíduos. A técnica consiste numa reação de PCR com dois primers marcados com fluorescência (HEX e FAM) correspondente ao alelo. A fluorescência de HEX é concebida para detectar um dos homozigotos; a fluorescência do FAM para detectar o outro; e fluorescência dentro do intermediário identificam o heterozigoto (SEMAGN et al., 2014).

Assim, a partir das técnicas de sequenciamento e genotipagem automatizadas a custo mais efetivo é possível a identificação de marcadores moleculares específicos e *screening* em grandes populações trazendo avanços consideráveis na SAM.

Associação genômica ampla

A associação genômica ampla (do inglês *Genome-wide Association Studies – GWAS*), tem ganhando espaço nos programas de melhoramento por permitir o mapeamento associativo em coleções de germoplasma ou populações naturais aproveitando de eventos de recombinação da população e do desequilíbrio de ligação para identificar mutações associadas a características fenotípicas (BURGHARDT; YOUNG; TIFFIN, 2017).

A partir de modelos robustos considerando o parentesco genômico e estrutura da população, para correção de efeitos poligênicos e pela correção por múltiplos testes (BUSH; MOORE, 2012) a GWAS pode fornecer informações sobre a arquitetura genética da característica, permitir a escolha dos parentais para mapeamento de *Quantitative Trait Loci* (QTL) e sugerir candidatos para mutagênese e transgenia (KORTE; FARLOW, 2013).

Em relação a cultura da mandioca os primeiros trabalhos de GWAS foram publicados em 2016 identificando SNPs relacionados a resistência ao vírus do mosaico africano (WOLFE et al., 2016) e relacionados ao conteúdo de carotenoides nas raízes (ESUMA et al., 2016). Posteriormente foram identificados SNPs relacionados ao conteúdo de matéria-seca nas raízes (RABBI et al., 2017), a podridão radicular (BRITO; OLIVEIRA; OLIVEIRA, 2017) e mais recentemente relacionado ao vírus do castanho listrado da mandioca (KAYONDO et al., 2018). A validação destes marcadores em conjunto maior de genótipos com a diversidade necessária proporcionará seu uso rotineiro nos programas de melhoramento da cultura. Após validação das variantes significativas identificadas pela GWAS, mutações causais em nível de nucleotídeo (*Quantitative Trait Nucleotide – QTNs*) podem ser utilizadas na Seleção Assistida por Marcadores moleculares (SAM) e/ou Seleção Genômica (GS) (KAMFWA; CICHY; KELLY, 2015) como ferramentas para a seleção precoce.

Características gerais da espectrometria no infravermelho próximo

A radiação infravermelha próxima (NIR) cobre a faixa do espectro eletromagnético entre os comprimentos de onda de 780 e 2500 nm. Sua utilização é baseada na correlação entre as propriedades de vibração diferencial das moléculas (dependente da sua estrutura molecular) e suas implicações com a absorção de luz em determinadas frequências (NICOLAÏ et al., 2007). Pode ser utilizada tanto para avaliação de característica quantitativas quanto qualitativas. Em abordagens qualitativas, como a utilizada no presente trabalho, métodos supervisionados ou não supervisionados para classificação ou predição podem ser adotados. No método supervisionado, o conjunto de espectros é usado para treinar o algoritmo de classificação relacionado a uma determinada classe enquanto nenhuma suposição a priori é realizada nos métodos não supervisionados (PASQUINI, 2003). Como exemplo podemos citar a identificação de espécies de plantas (CASTILLO et al., 2008; JIN et al., 2017; KOSMOWSKI; WORKU, 2018; LANG; ALMEIDA; COSTA, 2017), a discriminação de plantas transgênicas (LEE; CHOUNG, 2011; XIE et al., 2007) e a determinação da origem geográfica de variedades de plantas (CHEN; ZHAO; LIN, 2009; LI et al., 2012).

Outro ponto importante para operações com NIR é a fase de pré-processamentos dos espectros para remoção de ruídos. Alguns trabalhos avaliam inclusive a relação do pré-processamento dos espectros e sua capacidade de classificação/predição (MOSCETTI et al., 2013; XIN et al., 2012). Pré-processamentos comumente utilizados são (PASQUINI, 2003):

- média - utilizada para reduzir o número de comprimentos de onda ou suavizar o espectro;
- centralização - subtrai a média de cada variável e garante que todos os resultados sejam interpretáveis em termos de variação em torno da média, sendo recomendado para todas as aplicações práticas;
- normalização - remove o efeito de dispersão dos dados por diferentes métodos;
- suavização – remove o ruído geralmente pelo método de Savitzky–Golay.

Em relação a amostragem, a quantidade de amostras necessárias para produzir um modelo robusto deve incluir a variabilidade natural da concentração ou propriedade de interesse, de preferência uniformemente distribuídas entre os valores extremos (PASQUINI, 2003).

REFERÊNCIAS

- AIEMNAKA, P.; WONGKAEW, A.; CHANTHAWORN, J.; NAGASHIMA, S. K.; BOONMA, S.; AUTHAPUN, J.; JENWEERAWAT, S.; KONGSILA, P.; KITTIPADAKUL, P.; NAKASATHIEN, S.; SREEWONGCHAI, T.; WANNARAT, W.; VICHUKIT, V.; LÓPEZ-LAVALLE, L. A. B.; CEBALLOS, H.; ROJANARIDPICHED, C.; PHUMICHA, C. Molecular characterization of a spontaneous waxy starch mutation in cassava. **Crop Science**, v. 52, n. 5, p. 2121–2130, 2012.
- ASHOGBON, A. O.; AKINTAYO, E. T. Recent trend in the physical and chemical modification of starches from different botanical sources: A review. **Starch/Staerke**, v. 66, n. 1–2, p. 41–57, 2014.
- BALAGOPALAN, C. Cassava: biology, production and utilization. In: HILLOCKS, R. J.; THRESH, J. M. (Eds.). **Cassava: Biology, production and utilization**. Wallingford: CABI, 2002. p. 301–318.
- BARANDICA, O. J.; PÉREZ, J. C.; LENIS, J. I.; CALLE, F.; MORANTE, N.; PINO, L.; HERSHEY, C. H.; CEBALLOS, H. Cassava Breeding II: Phenotypic Correlations through the Different Stages of Selection. **Frontiers in Plant Science**, v. 7, p. 1–11, 2016.
- BREDESON, J. V.; LYONS, J. B.; PROCHNIK, S. E.; WU, G. A.; HA, C. M.; EDSINGER-GONZALES, E.; GRIMWOOD, J.; SCHMUTZ, J.; RABBI, I. Y.; EGESI, C.; NAULUVULA, P.; LEBOT, V.; NDUNGURU, J.; MKAMILO, G.; BART, R. S.; SETTER, T. L.; GLEADOW, R. M.; KULAKOW, P.; FERGUSON, M. E.; ROUNSLEY, S.; ROKHSAR, D. S. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. **Nature biotechnology**, v. 34, n. 5, p. 562–570, 2016.
- BREUNINGER, W. F.; PIYACHOMKWAN, K.; SRIROTH, K. Tapioca/Cassava Starch. In: **Starch**: Elsevier, 2009. p. 541–568.
- BRITO, A. C.; OLIVEIRA, S. A. S.; OLIVEIRA, E. J. Genome-wide association study for resistance to cassava root rot. **The Journal of Agricultural Science**, v. 155, n. 09, p. 1424–1441, 2017.
- BURGHARDT, L. T.; YOUNG, N. D.; TIFFIN, P. A Guide to Genome-Wide

Association Mapping in Plants. In: **Current Protocols in Plant Biology**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2017. p. 22–38.

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

CASTILLO, R.; CONTRERAS, D.; FREER, J.; RUIZ, J.; VALENZUELA, S. Supervised pattern recognition techniques for classification of Eucalyptus species from leaves NIR spectra. **Journal of the Chilean Chemical Society**, v. 53, n. 4, p. 1709–1713, 2008.

CEBALLOS, H.; CALLE, F. Cassava. In: JUAN, F.; NEBAMBI, L.; LARINDE, M.; BARKER, I.; ROCA, W.; CHUJOY, E. (Eds.). **Quality declared planting material Protocols and standards for vegetatively propagated crops**. Lima: Food & Agriculture Organization of the United Nations, 2007. p. 33–39.

CEBALLOS, H.; IGLESIAS, C. a.; PÉREZ, J. C.; DIXON, A. G. O. Cassava breeding: Opportunities and challenges. **Plant Molecular Biology**, v. 56, n. 4, p. 503–516, 2004.

CEBALLOS, H.; JARAMILLO, J. J.; SALAZAR, S.; PINEDA, L. M.; CALLE, F.; SETTER, T. Induction of flowering in cassava through grafting. **Journal of Plant Breeding and Crop Science**, v. 9, p. 19–29, 2017.

CEBALLOS, H.; KULAKOW, P.; HERSHEY, C. Cassava Breeding: Current Status, Bottlenecks and the Potential of Biotechnology Tools. **Tropical Plant Biology**, v. 5, n. 1, p. 73–87, 2012.

CEBALLOS, H.; SÁNCHEZ, T.; DENYER, K.; TOFIÑO, A. P.; ROSERO, E. A.; DUFOUR, D.; SMITH, A.; MORANTE, N.; PÉREZ, J. C.; FAHY, B. Induction and identification of a small-granule, high-amylose mutant in cassava (*Manihot esculenta* Crantz). **Journal of Agricultural and Food Chemistry**, v. 56, n. 16, p. 7215–7222, 2008.

CEBALLOS, H.; SÁNCHEZ, T.; MORANTE, N.; FREGENE, M.; DUFOUR, D.; SMITH, A. M.; DENYER, K.; PÉREZ, J. C.; CALLE, F.; MESTRES, C. Discovery of an amylose-free starch mutant in cassava (*Manihot esculenta* Crantz). **Journal of Agricultural and Food Chemistry**, v. 55, n. 18, p. 7469–7476, 2007.

CHEN, Q.; ZHAO, J.; LIN, H. Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 72, n. 4, p. 845–850, 2009.

CHUZEL, G. The Cassava Processing Industry in Brazil: Traditional Techniques, Technological Developments, Innovations and New Markets. **The African Journal of Food and Nutritional Security**, v. 1, n. 1, p. 46–59, 2001.

DE CARVALHO, R.; GUERRA, M. Cytogenetics of *Manihot esculenta* Crantz (cassava) and eight related species. **Hereditas**, v. 136, n. 2, p. 159–168, 2002.

EL-SHARKAWY, M. A. Physiological characteristics of cassava tolerance to prolonged drought in the tropics: implications for breeding cultivars adapted to seasonally dry and semiarid environments. **Brazilian Journal of Plant Physiology**, v. 19, n. 4, p. 257–286, 2007.

EL-SHARKAWY, M. A.; COCK, J. H.; PORTO, M. C. . Características fotossintéticas da mandioca (*Manihot esculenta* Crantz). **Revista Brasileira de Fisiologia Vegetal**, v. 1, n. 2, p. 143–154, 1989.

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**, v. 6, n. 5, p. 1–10, 2011.

ESUMA, W.; HERSELMAN, L.; LABUSCHAGNE, M. T.; RAMU, P.; 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.

FAOSTAT. **Food and Agriculture Organization of the United Nations**. 2018. Disponível em: <faostat.fao.org>.

FIETTO, J. L. R.; MACIEL, T. E. F. Sequenciando genomas. In: MOREIRA, L. M.; (Org). **Ciências genômicas: fundamentos e aplicações**. Riberão Preto: Sociedade Brasileira de Genética, 2015. p. 27–64.

FOOD INGREDIENTS BRASIL. Amidos estrutura e fontes. **Food Ingredients Brasil**, v. 35, p. 1–26, 2015.

FUKUDA, W. N. G.; GUEVARA, C. L. **Descritores morfológicos e agronômicos para a caracterização de mandioca (*Manihot esculenta* Crantz)**, Documentos 78, EMBRAPA-CNPMP, 1998.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Levantamento Sistemático da Produção Agrícola**. 2018. Disponível em: <<https://sidra.ibge.gov.br/home/lspa/brasil>>.

JENNINGS, D. L.; IGLESIAS, C. Plant Breeding and Crop Improvement. In: HILLOCKS, RORY J.; THRESH, J. M.; BELLOTTI, A. (Ed.). **Cassava: Biology, production and utilization**. Chichester, UK: CABI, 2002. p. 149–166.

JIN, X.; CHEN, X.; XIAO, L.; SHI, C.; CHEN, L.; YU, B.; YI, Z.; YOO, J. H.; HEO, K.; YU, C. Y.; YAMADA, T.; SACKS, E. J.; PENG, J. Application of visible and near-infrared spectroscopy to classification of *Miscanthus* species. **PLoS ONE**, v. 12, n. 4, p. 1–12, 2017.

KAMFWA, K.; CICHY, K. A.; KELLY, J. D. Genome-Wide Association Study of Agronomic Traits in Common Bean. **The Plant Genome**, v. 8, n. 2, p. 1–12, 2015.

KAYONDO, S. I.; PINO DEL CARPIO, D.; LOZANO, R.; OZIMATI, A.; WOLFE, M.; BAGUMA, Y.; GRACEN, V.; OFFEI, S.; FERGUSON, M.; KAWUKI, R.; JANNINK, J.-L. Genome-wide association mapping and genomic prediction for CBSD resistance in *Manihot esculenta*. **Scientific Reports**, v. 8, n. 1, p. 1549, 2018.

KEELING, P. L.; MYERS, A. M. Biochemistry and Genetics of Starch Synthesis. **Annual Review of Food Science and Technology**, v. 1, n. 1, p. 271–303, 2010.

KORTE, A.; FARLOW, A. The advantages and limitations of trait analysis with GWAS: a review. **Plant Methods**, v. 9, n. 1, p. 29, 2013.

KOSMOWSKI, F.; WORKU, T. Evaluation of a miniaturized NIR spectrometer for cultivar identification: The case of barley, chickpea and sorghum in Ethiopia. **PLoS ONE**, v. 13, n. 3, p. 1–17, 2018.

LANG, C.; ALMEIDA, D. R. A.; COSTA, F. R. C. Discrimination of taxonomic identity at species, genus and family levels using Fourier Transformed Near-Infrared Spectroscopy (FT-NIR). **Forest Ecology and Management**, v. 406, p. 219–227, 2017.

LATIF, S.; MÜLLER, J. Potential of cassava leaves in human nutrition: A review. **Trends in Food Science and Technology**, v. 44, n. 2, p. 147–158, 2015.

LEE, J. H.; CHOUNG, M.-G. Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. **Food Chemistry**, v. 126, n. 1, p. 368–373, 2011.

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

LI, B.; WEI, Y.; DUAN, H.; XI, L.; WU, X. Discrimination of the geographical origin of *Codonopsis pilosula* using near infrared diffuse reflection spectroscopy coupled with random forests and k-nearest neighbor methods. **Vibrational Spectroscopy**, v. 62, p. 17–22, 2012.

MOORTHY, S. N. Physicochemical and Functional Properties of Tropical Tuber Starches: A Review. **Starch - Stärke**, v. 54, n. 12, p. 559–592, 2002.

MOSCETTI, R.; HAFF, R. P.; AERNOUTS, B.; SAEYS, W.; MONARCA, D.; CECCHINI, M.; MASSANTINI, R. Feasibility of Vis/NIR spectroscopy for detection of flaws in hazelnut kernels. **Journal of Food Engineering**, v. 118, n. 1, p. 1–7, 2013.

NG, L. M.; SIMMONS, R. Infrared Spectroscopy. In: **Modern Vibrational Spectroscopy and Micro-Spectroscopy**. Chichester, UK: John Wiley & Sons, Ltd, 2015. v. 71p. 71–102.

NICOLAÏ, B. M.; BEULLENS, K.; BOBELYN, E.; PEIRS, A.; SAEYS, W.; THERON, K. I.; LAMMERTYN, J. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. **Postharvest Biology and Technology**, v. 46, n. 2, p. 99–118, 2007.

OLIVEIRA, A. C. De; DA MAIA, L. C.; FARIAS, D. da R.; MARINI, N. Genômica. In: BORÉM, A.; FRITSCHÉ-NETO, R. (Eds.). **Ômicas 360° - Aplicações e estratégias para o melhoramento de plantas**. Viçosa: UFV, 2013. p. 23–46.

OLSEN, K. M.; SCHAAL, B. a. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. **Proceedings of the National Academy of Sciences of the**

United States of America, v. 96, p. 5586–5591, 1999.

PASQUINI, C. Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198–219, 2003.

PFISTER, B.; ZEEMAN, S. C. Formation of starch in plant cells. **Cellular and Molecular Life Sciences**, v. 73, n. 14, p. 2781–2807, 2016.

PROCHNIK, S.; MARRI, P. R.; DESANY, B.; RABINOWICZ, P. D.; KODIRA, C.; MOHIUDDIN, M.; RODRIGUEZ, F.; FAUQUET, C.; TOHME, J.; HARKINS, T.; ROKHSAR, D. S.; ROUNSLEY, S. The Cassava Genome: Current Progress, Future Directions. **Tropical Plant Biology**, v. 5, n. 1, p. 88–94, 2012.

RABBI, I. Y.; UDOH, L. I.; WOLFE, M.; PARKES, E. Y.; GEDIL, M. A.; DIXON, A.; RAMU, P.; JANNINK, J.-L.; KULAKOW, P. Genome-Wide Association Mapping of Correlated Traits in Cassava: Dry Matter and Total Carotenoid Content. **The Plant Genome**, v. 10, n. 3, p. 1–14, 2017.

RAHEEM, D.; CHUKWUMA, C. Foods from cassava and their relevance to Nigeria and other African countries. **Agriculture and Human Values**, p. 383–390, 2001.

REINCKE, K.; VILVERT, E.; FASSE, A.; GRAEF, F.; SIEBER, S.; LANA, M. A. Key factors influencing food security of smallholder farmers in Tanzania and the role of cassava as a strategic crop. **Food Security**, p. 1–14, 2018.

SAITHONG, T.; RONGSIRIKUL, O.; KALAPANULAK, S.; CHIEWCHANKASET, P.; SIRIWAT, W.; NETRPHAN, S.; SUKSANGPANOMRUNG, M.; MEECHAI, A.; CHEEVADHANARAK, S. Starch biosynthesis in cassava: a genome-based pathway reconstruction and its exploitation in data integration. **BMC Systems Biology**, v. 7, n. 1, p. 2–17, 2013.

SÁNCHEZ, T.; DUFOUR, D.; MORENO, I. X.; CEBALLOS, H. Comparison of pasting and gel stabilities of waxy and normal starches from potato, maize, and rice with those of a novel waxy cassava starch under thermal, chemical, and mechanical stress. **Journal of Agricultural and Food Chemistry**, v. 58, n. 8, p. 5093–5099, 2010.

SANCHEZ, T.; SALCEDO, E.; CEBALLOS, H.; DUFOUR, D.; MAFLA, G.;

MORANTE, N.; CALLE, F.; PÉREZ, J. C.; DEBOUCK, D.; JARAMILLO, G.; XIMENA MORENO, I. Screening of starch quality traits in cassava (*Manihot esculenta* Crantz). **Starch - Stärke**, v. 61, n. 1, p. 12–19, 2009.

SEMAGN, K.; BABU, R.; HEARNE, S.; OLSEN, M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. **Molecular Breeding**, v. 33, n. 1, p. 1–14, 2014.

SINGH, N.; SINGH, J.; KAUR, L.; SODHI, N. S.; GILL, B. S. Morphological, thermal and rheological properties of starches from different botanical sources. **Food Chemistry**, v. 81, n. 2, p. 219–231, 2003.

TAIZ, L.; ZEIGER, E. **Fisiologia e desenvolvimento vegetal**. 6. ed. Porto Alegre: Artmed, 2017.

VENTURINI, M. T.; SANTOS, L. R.; VILDOSO, C. I. A.; SANTOS, V. S.; OLIVEIRA, E. J. Variation in cassava germplasm for tolerance to post-harvest physiological deterioration. **Genetics and Molecular Research**, v. 15, n. 2, p. 1–18, 2016.

WOLFE, M. D.; RABBI, I. Y.; EGESI, C.; HAMBLIN, M.; KAWUKI, R.; KULAKOW, P.; LOZANO, R.; CARPIO, D. P. Del; RAMU, P.; JANNINK, J.-L. Genome-Wide Association and Prediction Reveals Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement. **The Plant Genome**, v. 9, n. 2, 2016.

XIE, L.; YING, Y.; YING, T.; YU, H.; FU, X. Discrimination of transgenic tomatoes based on visible/near-infrared spectra. **Analytica Chimica Acta**, v. 584, n. 2, p. 379–384, 2007.

XIN, N.; GU, X. F.; WU, H.; HU, Y. Z.; YANG, Z. L. Discrimination of raw and processed *Dipsacus asperoides* by near infrared spectroscopy combined with least squares-support vector machine and random forests. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 89, p. 18–24, 2012.

ARTIGO 1

IDENTIFICAÇÃO E VALIDAÇÃO DE MUTAÇÕES PONTUAIS ASSOCIADAS COM O FENÓTIPO WAXY EM MANDIOCA¹

¹Artigo a ser ajustado para posterior submissão ao Comitê Editorial do periódico científico *Plant Molecular Biology*, em versão na língua inglesa.

Identificação e validação de mutações pontuais associadas com o fenótipo *waxy* em mandioca

Resumo: A enzima *Granule-bound starch synthase I (GBSSI)* é responsável pela síntese da amilose, e, portanto, sua ausência resulta em indivíduos com fenótipo ceroso (*waxy*) em diversas culturas amiláceas. Este trabalho teve como objetivo validar mutações pontuais previamente associadas com o fenótipo *waxy* no germoplasma de mandioca da América Latina, bem como identificar outros alelos mutantes alternativos no gene *GBSSI* que possam ser úteis na seleção assistida para esta característica. Na primeira etapa, 1529 acessos de mandioca foram genotipados com *Single Nucleotide Amplified Polimorfism (SNAP)* associados ao amido ceroso. Destes, 206 acessos apresentaram os alelos mutantes em heterozigose, dos quais 28 foram autofecundados para geração da S₁. A segregação molecular de três das 28 famílias autofecundadas seguiu a proporção 3:1 (gene dominante para presença de amilose), sendo identificados 41 indivíduos recessivos, mas que não expressaram o fenótipo *waxy*. Na segunda etapa, o gene *GBSSI* foi sequenciado em 31 indivíduos *waxy* e 59 não-*waxy*, objetivando identificar alelos mutantes alternativos. Como resultado, foram identificadas 17 mutações no gene *GBSSI* (16 SNPs e uma deleção), das quais apenas a deleção no éxon 6 (MeWxEx6-del-C) foi correlacionada ao fenótipo *waxy* (1701pb, deleção de citosina). A avaliação dos pontos de mutação por análise discriminante de componentes principais (ADCP) também não discriminou completamente os indivíduos cerosos. A probabilidade de designação de indivíduos com base no genótipo associado a cada tipo de amido diferiu da classificação fenotípica. Na terceira etapa os acessos de mandioca foram genotipados via KASP (*Kompetitive Allele Specific PCR*), cujos resultados demonstraram a inexistência de indivíduos heterozigóticos para a deleção MeWxEx6-del-C no germoplasma analisado. Portanto, a deleção MeWxEx6-del-C não deve ser utilizada para seleção assistida em backgrounds genéticos diferentes da fonte original do amido *waxy*. Além disso, os SNPs alternativos identificados neste trabalho também não foram associados ao fenótipo *waxy* quando se analisou um painel de acessos com alta diversidade genética.

Palavras-chave: amido ceroso; *screening*; sequenciamento; KASP.

Identification and validation of mutation points associated with waxy phenotype in cassava

Abstract: The enzyme Granule-bound starch synthase I (GBSSI) is responsible for the synthesis of amylose and, therefore, its absence results in individuals with waxy phenotype in various amylaceous crops. This study aimed to validate mutation points previously associated with the waxy phenotype in the cassava germplasm of Latin America, as well as to identify other alternative mutant alleles in the GBSSI gene that may be useful in assisted selection for this trait. In the first step, 1529 cassava accessions were genotyped with Single Nucleotide Amplified Polymorphism (SNAP) associated with waxy starch. 206 accessions had mutant alleles in heterozygosis, from which 28 were self-fertilized for the S₁ generation. The molecular segregation from three of the 28 self-fertilized families followed the 3:1 ratio (dominant gene for the presence of amylose). 41 recessive individuals were identified, but none of those expressed the waxy phenotype. In the second step, the GBSSI gene was sequenced in 31 waxy and 59 non-waxy individuals, aiming to identify alternative mutant alleles. As a result, 17 mutations in the GBSSI gene (16 SNPs and one deletion) were identified, in which only the deletion in exon 6 (MeWxEx6-del-C) was correlated to the waxy phenotype (1701bp, cytokine deletion). The evaluation of mutation points by discriminant principal component analysis (DAPC) also did not completely discriminate the waxy individuals. The probability of attribution based on the genotypic information for each type of starch differed from the phenotypic classification. In the third step, the accessions of cassava were genotyped by Kompetitive Allele Specific PCR (KASP). The results demonstrated the non-existence of heterozygous individuals for the MeWxEx6-del-C deletion in the analyzed germplasm. Therefore, the deletion MeWxEx6-del-C should not be used for assisted selection in genetic backgrounds different from the original source of waxy starch. Also, the alternative SNPs identified in this study were not associated with the waxy phenotype when compared to a panel of accessions with high genetic diversity.

Keywords: waxy starch; screening; sequencing; KASP.

INTRODUÇÃO

O amido é amplamente utilizado em múltiplos fins comerciais, sobretudo por fornecer mais de 80% das calorias utilizadas na dieta humana (KEELING; MYERS, 2010). É basicamente constituído por dois tipos de polímeros: amilose (essencialmente α 1,4-poliglucanos - lineares) e amilopectina (α 1,4-poliglucanos e α -1,6-poliglucanos - ramificados) (PÉREZ; BERTOFT, 2010) cujas proporções e arranjos conferem características diferenciadas que definem seu uso e são foco dos programas de melhoramento genético. As principais fontes comerciais de amido são: milho, arroz, trigo, batata e mandioca, cujas propriedades funcionais e específicas de cada uma destas espécies definem suas aplicações industriais (COPELAND et al., 2009; HOOVER, 2001; SINGH et al., 2003; WANG; HENRY; GILBERT, 2014).

A amilopectina é a unidade básica do grânulo de amido composta de muitas cadeias curtas de moléculas de glicose. Em contrapartida, a amilose que é uma molécula menor com cadeias mais longas (DENYER et al., 2001). O amido normalmente possui entre 20 a 30% de amilose e 70 a 80% de amilopectina (KEELING; MYERS, 2010). Amidos que possuem baixa amilose, conhecidos como tipo ceroso (*waxy*) são importantes por apresentar menor sinérese e retrogradação, serem claros e apresentarem um gel mais viscoelástico (ŠÁRKA; DVOŘÁČEK, 2017). A retrogradação é um processo no qual as cadeias de amilose e amilopectina desagregadas em uma pasta aquecida tendem a sofrer um processo de cristalização. Com o tempo esta reorganização pode liberar água retida dentro da estrutura (sinérese) com efeito prejudicial na qualidade sensorial e de armazenamento nos alimentos derivados deste tipo de amido, sobretudo porque altera as propriedades texturais e nutricionais dos alimentos (WANG et al., 2015). Além do amido ceroso de mandioca apresentar menor sinérese quando comparado ao amido ceroso de outras culturas (SÁNCHEZ et al., 2010) o amido de mandioca possui sabor neutro devido ao seu baixo teor de lipídios e proteínas o que lhe confere vantagens competitivas em relação aos amidos de cereais para uso na indústria de alimentos (WATERSHOOT et al., 2015).

Mesmo havendo a possibilidade de uso de modificações químicas e físicas do amido para alterar suas propriedades funcionais para atender a demandas específicas de mercado, existem certas limitações nestas modificações para

algumas finalidades. Além disso, as modificações no amido, independentemente do tipo, sempre estão associadas a um maior custo industrial. Outro aspecto, a ser considerado é que os consumidores estão cada vez mais exigindo produtos que sejam mais naturais, como mínimo de intervenção industrial. Portanto, existe um amplo mercado para desenvolvimento e exploração de amidos naturalmente diferenciados. Assim, os programas de melhoramento genético podem contribuir na descoberta de variantes genéticas de interesse e incorporação destas características em variedades comerciais, e o amido ceroso é um destes amidos especiais.

Do ponto de vista genético, a exploração das propriedades dos amidos deve ser precedida da compreensão dos mecanismos moleculares relacionados à sua biossíntese. Dentre as enzimas relacionadas à síntese de amido, a *GBSSI* (*Granule-bound starch synthase I*) é uma sintase ligada ao alongamento da amilose e mutantes cerosos de muitas espécies apresentam atividade deficiente dessa enzima (DENYER et al., 2001; TETLOW; MORELL; EMES, 2004; ZEEMAN; KOSSMANN; SMITH, 2010). Diferentes formas alélicas derivadas de diferentes mecanismos moleculares foram identificadas como responsáveis pelo fenótipo *waxy*. Em arroz, dois SNPs foram identificados no éxon 6 e 10 do gene *GBSSI* com associação ao conteúdo de amilose e viscosidade da pasta (LARKIN; PARK, 2003). Uma forte correlação também foi observada entre um SNP localizado no éxon1/intron1 com as propriedades de pasta em amido de arroz (KHARABIAN-MASOULEH et al., 2012). No trigo hexaplóide comum (*Triticum aestivum* L.), cada um dos genomas A, B e D tem uma proteína cerosa, cujos respectivos *locus* codificadores (*Wx-A1*, *Wx-B1*, e *Wx-D1*) possuem várias mutações relatadas na literatura. Em outro trabalho a inserção de um SNP e uma deleção de um único nucleotídeo no alelo nulo *Wx-A1* induziram códons de terminação prematuros a cerca de 55 nucleotídeos a frente do ponto de mutação em *Triticum dicoccoides* A. e *T. dicoccum* S (SAITO; NAKAMURA, 2005). Mais recentemente um novo alelo nulo foi caracterizado pela inserção de um transposon e consequente perda de função nesta espécie (ZHANG et al., 2017). Duas novas mutações também foram recentemente descobertas em milho relacionadas a elementos transponíveis, sendo um retrotransposon de 466-bp inserido no éxon seis e um elemento transponível de repetição invertida de 116 pb inserido no éxon sete (XIAOYANG et al., 2017).

Em mandioca, estudos sobre o controle genético do fenótipo ceroso e mesmo sua exploração para o desenvolvimento de variedades com esta característica são relativamente recentes. Os primeiros mutantes cerosos em mandioca foram induzidos por modificações na expressão do gene *GBSSI* via transgenia por Raemakers et al. (2005) e posteriormente por Zhao et al. (2011) e Koehorst-Van Putten et al. (2012). Em seguida, uma mutação natural foi relatada por Ceballos et al. (2007) numa série de autofecundações realizadas no Centro Internacional de Agricultura Tropical – CIAT que evidenciou a condição recessiva da característica derivada do genótipo AM206-5.

Aiennaka et al. (2012) realizaram uma caracterização molecular do gene *GBSSI* na fonte AM206-5 e identificaram mutações funcionais potenciais. Os autores encontraram um InDel no éxon 6 com uma única exclusão de base (citosina) que cria um códon de parada prematuro; uma variante de duas bases no éxon 11 (GC e AT); e uma substituição no íntron 11 (C para G). Com esta última mutação os autores desenvolveram marcadores SNAP (*Single-Nucleotide-Amplified Polymorphism*) cujas informações permitiriam direcionar autofecundações e o retrocruzamento assistido por marcadores moleculares reduzindo custos e tempo. Assim, o presente trabalho teve como objetivos: *i*) validar mutações (SNPs e InDels) previamente associadas a fonte original do amido ceroso AM206-5 no germoplasma de mandioca da América Latina; e *ii*) identificar alelos alternativos no gene *GBSSI* em um painel de acessos *waxy* e não-*waxy* que possam ser úteis na seleção assistida para esta característica.

MATERIAL E MÉTODOS

Screening do germoplasma de mandioca com marcadores SNAP (Single Nucleotide Amplified Polimorfism) associados ao fenótipo waxy na fonte AM206-5

No total foram analisados 1529 acessos pertencentes ao Banco Ativo de Germoplasma de Mandioca (BAG-Mandioca) da Embrapa Mandioca e Fruticultura (CNPMPF) procedentes de vários ecossistemas do Brasil, Colômbia, Venezuela e Uganda. A extração do DNA foi realizada conforme o protocolo CTAB (brometo de cetiltrimetilamônio) (DOYLE; DOYLE, 1990). Com o intuito de verificar a qualidade e quantidade do DNA extraído, realizou-se quantificação em gel de agarose 1%

(p/v) (Invitrogen, EUA) corado com brometo de etídio ($1,0 \text{ mg.L}^{-1}$) por comparação visual com diversas concentrações de DNA fago Lambda (Invitrogen, EUA).

As mutações localizadas no éxon 6 (deleção de 1 pb no nucleotídeo 92) e o SNP no éxon 11 (GC para AT) encontradas por Aiemnaka et al. (2012) não foram otimizadas mesmo após diversos ajustes nas reações de PCR. A genotipagem dos acessos de mandioca foi então realizada utilizando apenas os iniciadores SNAP (*Single Nucleotide Amplified Polimorfism*) MeWxl11-G e MeWxl11-C desenvolvidos por Aiemnaka et al. (2012) localizado no íntron 11 do gene *waxy* (GBSSI). As reações de *Polymerase Chain Reaction* (PCR) foram realizadas em volume final de $15 \mu\text{L}$ contendo 10 ng de DNA, tampão de PCR 1X, 1,5 mM de MgCl_2 (4G, Brasil), 0,2 mM de dNTP (Promega, EUA), 0,2 μM de cada iniciador (Integrated DNA Technologies, EUA) e 1U de Taq DNA Polimerase, obtida conforme Pluthero (1993). O programa de amplificação consistiu de um ciclo a 94°C por 2 min; 30 ciclos a 94°C por 30s, temperatura de anelamento de 56°C por 30s, 72°C por 1 min; e extensão final a 72°C por 5 min em termociclador modelo Veriti® 96-well (Applied Biosystems, EUA). Para a revelação dos produtos de amplificação, utilizou-se gel de agarose 1000 2% (p/v) (Invitrogen, EUA) contendo $1,0 \text{ mg.L}^{-1}$ de brometo de etídio e submetida a eletroforese em tampão TBE 0,5 x (45 mM Tris-borate, 1 mM EDTA e q.s.p de água destilada). Os produtos no gel foram visualizados em luz UV e registrados com o fotodocumentador Gel Logic 212 Pro (*Carestream Molecular Imaging*, EUA).

Avaliação de progênies S_1 derivadas do screening com marcadores SNAP

A partir da genotipagem realizada no BAG-Mandioca, 28 acessos previamente identificados contendo o alelo G na posição 3197pb do íntron 11 (MeWxl11-G) na forma heterozigótica e associado ao fenótipo *waxy* (AIEMNAKA et al., 2012) foram autofecundados para geração de indivíduos homozigóticos com fenótipo *waxy*. Os acessos selecionados para autofecundação de acordo com seu florescimento no período de agosto a novembro de 2014 foram: BGM0061, BGM0131, BGM0132, BGM0222, BGM0463, BGM0505, BGM0614, BGM0650, BGM0726, BGM0729, BGM0741, BGM0872, BGM0935, BGM0941, BGM0962, BGM1023, BGM1041, BGM1120, BGM1143, BGM1148, BGM1253, BGM1284, BGM1288, BGM1335, BGM1378, BGM1383, BGM1413 e BGM1819.

Para realizar a autofecundação dos acessos, as flores femininas receptivas no dia foram cobertas com sacos de pano (20 x 15 cm) de manhã antes da abertura, para protegê-los da contaminação. Flores masculinas do mesmo genótipo foram coletadas pela manhã e colocadas em frascos de tampa larga, previamente rotulados. No final da manhã e início da tarde, foram realizadas as autofecundações por meio do contato das anteras com o estigma da flor feminina para garantir a polinização. As flores polinizadas foram cobertas com um voil (tecido de algodão leve) até a coleta das sementes após deiscência natural.

As sementes oriundas da autofecundação dos 28 acessos foram semeadas em casa de vegetação e após 30 dias transplantadas para o campo. A colheita e avaliação das plantas foram realizadas dez meses após o plantio. As plantas S₁ de cada uma das famílias foram avaliadas quanto à presença do amido ceroso (*waxy*) por meio do teste indicativo de iodo a 2% (2g KI e 0.2g I₂ em água destilada), aplicado em uma secção transversal das raízes de todos os genótipos (CEBALLOS et al., 2007). As longas cadeias de amilose têm uma alta capacidade de se ligar à ionina em solução e isso confere uma cor azul aos amidos contendo amilose quando coradas com iodo. Em contrapartida, a amilopectina tem uma baixa capacidade de ligação ao iodo e por isso manchas vermelho-marrom são características de amido contendo essencialmente este polímero (DENYER et al., 2001).

Após as avaliações das famílias S₁ no campo, 185 genótipos pertencentes às famílias S₁-BGM0061 (60 indivíduos); S₁-BGM0463 (24 indivíduos) e S₁-BGM0935 (101 indivíduos) foram genotipados com marcadores SNAP, conforme descrito anteriormente. A segregação molecular para o gene *GBSSI* nos indivíduos S₁ com base nos iniciadores MeWxl11-G e MeWxl11-C foi avaliada pelo teste de Qui-quadrado $\sum_{i=1}^n \frac{(f_o - f_e)^2}{f_e}$, em que f_o e f_e são as frequências observadas e esperadas dos fenótipos.

Sequenciamento do gene *GBSSI*

Com o intuito de identificar novas variantes alélicas associadas ao amido ceroso foi realizado o sequenciamento completo do gene *GBSSI* em 89 genótipos de mandioca, sendo 54 acessos de germoplasma no estado homozigótico (CC) ou heterozigótico (CG) identificados pelos iniciadores MeWxl11-G e MeWxl11-C

(fenótipo não-*waxy*); e 35 genótipos derivados de populações segregantes com a fonte AM206-5, sendo 3 homozigóticos (CC) ou heterozigóticos (CG) (fenótipo não-*waxy*) e 32 genótipos homozigóticos (GG) (fenótipo *waxy*) (Tabela 1). A extração e a quantificação do DNA genômico foi realizada conforme descrito anteriormente.

Tabela 1: Relação de acessos de mandioca utilizados no sequenciamento do gene *GBSSI* (*Granule-bound starch synthase I*).

Acessos não-<i>waxy</i> (<i>Wx</i>_)				
9602-02	BGM0171	BGM0729	BGM1198	BGM1453
9607-07	BGM0222	BGM0741	BGM1287	BGM1594
BGM0023	BGM0263	BGM0752	BGM1288	BGM1596
BGM0043	BGM0326	BGM0756	BGM1328	BGM1598
BGM0046	BGM0336	BGM0776	BGM1335	BGM1692
BGM0087	BGM0360	BGM0785	BGM1339	BGM1711
BGM0116	BGM0362	BGM0841	BGM1342	7734-03
BGM0132	BGM0378	BGM1118	BGM1354	7734-05
BGM0145	BGM0399	BGM1140	BGM1364	7745-01
BGM0146	BGM0536	BGM1143	BGM1444	
BGM0149	BGM0545	BGM1144	BGM1448	
BGM0170	BGM0592	BGM1161	BGM1451	
Acessos <i>Waxy</i> (<i>wxwx</i>)				
6383-01	7690-08	7734-06	7737-07	7747-01
6413-01	7694-03	7734-07	7737-08	8109-03
6421-05	7697-03	7736-05	7738-01	8109-04
6428-01	7698-14	7736-07	7745-02	8109-05
6430-01	7698-16	7737-01	7745-04	
6437-01	7734-01	7737-02	7745-05	
6477-02	7734-02	7737-06	7745-06	

Para a amplificação completa do gene foram desenvolvidos cinco iniciadores tendo como sequência referência o gene *GBSSI* (Manes.02G001000) do genoma de mandioca v6.1 (BREDESON et al., 2016) depositado no banco de dados Phytozome v12.1 (GOODSTEIN et al., 2012) (Tabela 2). O programa Primer3 (UNTERGASSER et al., 2012) foi utilizado para desenhar os iniciadores com os seguintes critérios: tamanho do produto amplificado entre 800 e 1100pb; temperatura de anelamento acima de 60° C; e porcentagem de G/C acima de 40%. Os iniciadores foram alocados de forma a sobrepor cada uma das sequências e garantir cobertura completa do gene.

Tabela 2: Relação de iniciadores utilizados para amplificação e sequenciamento completo do gene *GBSSI* (*Granule-bound starch synthase I*) em genótipos *waxy* e não-*waxy* de mandioca.

Iniciador	Sequência	Tamanho esperado do fragmento (pb)
MeGBSSIpA-F	TGGCGAAGTCCCACCATTAC	974
MeGBSSIpA-R	TGTACTGGTCATAGCGGGGA	
MeGBSSIpB-F	TCCCCGCTATGACCAGTACA	988
MeGBSSIpB-R	ACAAGTCACCAACCCCGAAA	
MeGBSSIpC-F	CACTGCTCTGCTTCCATGTTATCT	936
MeGBSSIpC-R	TCTCACAACACAACCAAGGACATC	
MeGBSSIpD-F	CAGAAAGTCGGATTGCCTGTTGATA	931
MeGBSSIpD-R	GCTACCAGTCAATCCAATTTGCAC	
MeGBSSIpE-F	TGACTAAGTATCTAGGAGGCTCA	1038
MeGBSSIpE-R	GAAGGGAAGAAAGAACTGAATGAC	

Para otimização dos iniciadores foram testadas diferentes temperaturas de anelamento (58 a 64° C), diferentes concentrações de cloreto de magnésio (1, 1,5 e 2mM) e diferentes números de ciclos de extensão da PCR. As reações de PCR foram otimizados em volume final de 50 µL contendo 10 ng de DNA, tampão de PCR 1X, 1,5/2 mM de MgCl₂ (Invitrogen, EUA), 0,2 mM de dNTP (Promega, EUA), 0,2 mM de cada iniciador (Integrated DNA Technologies, EUA) e 1U de Taq DNA Polimerase High fidelity (Invitrogen, EUA). O programa de amplificação foi otimizado utilizando uma desnaturação inicial de um ciclo a 95 °C por 1 min; seguido de 30/35 ciclos a 95 °C por 15s, temperatura de anelamento 62° C por 15s, 72 °C por 30s; e extensão final a 72°C por 7 min em termociclador modelo Veriti® 96-well (Applied Biosystems, EUA).

O produto da eletroforese foi purificado com ExoSap-IT (Affymetrix, EUA), sequenciado em ambas direções usando BigDye® Terminator v3.1 Cycle Sequencing (Applied Biosystems, EUA) e analisado em sequenciador ABI PRISM 3730 XL Analyzer (Applied Biosystems, EUA). Os cromatogramas foram atribuídos e trimados com Phred (EWING et al., 1998; EWING; GREEN, 1998), alinhados com a sequência referência Manes.02G001000 e os SNPs identificados com o auxílio do programa novoSNP v3.0.1 (WECKX et al., 2005). A identificação das regiões

codificadoras e não codificadoras foi realizada por comparação com o gene *GBSSI* (Manes.02G001000) do genoma de mandioca v6.1 (BREDESON et al., 2016).

Análise discriminante de componentes principais

A análise discriminante de componentes principais (ADCP) foi realizada para avaliar o poder de discriminação do fenótipo *waxy* com base nas variantes alélicas identificadas pelo sequenciamento completo do gene *GBSSI*. A ADCP foi realizada retendo mais de 90% da variação dos dados com os 10 primeiros componentes principais e uma função discriminante. A análise foi realizada no ambiente de programação R v3.5 (R CORE TEAM, 2018) com auxílio do pacote *adegenet* (JOMBART; AHMED, 2011)

Avaliação genotípica do banco de germoplasma para deleção MeWxEx6-del-C

A partir dos resultados do sequenciamento completo do gene *GBSSI* identificou-se uma deleção no éxon 6 na posição 1701pb (daqui em diante denominada MeWxEx6-del-C), já descrita por Aiemnaka et al. (2012). Esta deleção foi utilizada para o *screening* do germoplasma de mandioca utilizando a tecnologia *Kompetitive Allele Specific PCR* (KASP) (SEMAGN et al., 2014) pela empresa Intertek AgriTech. Na genotipagem anterior os iniciadores SNAP MeWxl11-G e MeWxl11-C foram enriquecidos com três nucleotídeos na extremidade 3' para permitir a amplificação diferencial de um SNP no íntron 11. A mesma estratégia foi utilizada nas demais mutações reportadas por Aiemnaka et al. (2012) sem sucesso (dados não apresentados). Em contrapartida a tecnologia KASP consegue detectar uma base de diferença pelo uso de iniciadores diferentemente marcados por fluorescência (SEMAGN et al., 2014).

A otimização da genotipagem via KASP foi realizada em duas etapas, sendo que na primeira etapa foi utilizado um conjunto de indivíduos com condição alélica conhecida, ou seja 31 indivíduos homozigóticos para o gene *waxy* (*wxwx*); 31 indivíduos heterozigóticos (*Wxwx*) oriundos de populações segregantes para o genótipo *waxy* derivado da fonte AM206-5 e 32 indivíduos não-*waxy* (*WxWx*) do BAG-Mandioca (Tabela 3). A fluorescência específica do alelo (HEX e FAM) foi detectada usando o programa SNPviewer2 v4.0.0. e agrupamentos visualizados com auxílio do ambiente de programação R v.3.5 (R CORE TEAM, 2018). Após

otimização e validação do marcador MeWxEx6-del-C foi realizada a genotipagem de 1529 acessos do BAG-Mandioca.

Tabela 3: Genótipos de mandioca utilizadas para otimização da genotipagem da deleção de uma citosina no éxon 6, posição 1701pb (MeWxEx6-del-C) via *Kompetitive Allele Specific PCR* (KASP).

Amostras waxy homozigóticas (wxwx)				
6460-2	7474-1	7799-2	7909-6	8034-2
6466-3	7738-4	7802-3	7921-1	8093-3
6502-1	7745-5	7807-5	7934-1	8109-4
6703-1	7751-1	7811-6	7950-3	
6896-4	7754-3	7813-1	7953-2	
7020-1	7773-6	7867-4	7992-3	
7429-3	7788-7	7882-2	8014-7	
Amostras não-waxy heterozigóticas (Wxwx)				
2017wx-01-01	2017wx-02-12	2017wx-02-43	2017wx-03-09	2017wx-03-29
2017wx-01-02	2017wx-02-17	2017wx-03-01	2017wx-03-16	2017wx-03-31
2017wx-01-03	2017wx-02-18	2017wx-03-02	2017wx-03-18	2017wx-03-38
2017wx-01-06	2017wx-02-19	2017wx-03-03	2017wx-03-20	
2017wx-01-07	2017wx-02-25	2017wx-03-06	2017wx-03-21	
2017wx-02-10	2017wx-02-26	2017wx-03-07	2017wx-03-22	
2017wx-02-11	2017wx-02-36	2017wx-03-08	2017wx-03-28	
Amostras não-waxy homozigóticas (WxWx)				
032-09	BGM2326	Conquista 1	JoselitoA2	Roxona
517-08	BGM2327	Conquista 2	Ouro Pão	RR0065
Aciolina	BGM2333	CS01	Peru Preto	Tailandesa
Aipim Abacate	BRS396	Folha Fina	Pretinha	Venâncio-RN
Amarelona	BRS399	Inajazinha	RI-F	
AM-Jaeve-RN	CL-Acre	Ipirá	Retori	
BGM0001	CL-RI	Jacona	Roxinha	

RESULTADOS

Screening do germoplasma de mandioca com marcadores SNAP

Para a genotipagem do germoplasma de mandioca os iniciadores SNAP relacionados ao íntron 11, gerado pela substituição da base C por G desenvolvidos por Aiemnaka et al. (2012) foram otimizados e utilizados no *screening* de alelos associados ao fenótipo *waxy* (G) no germoplasma de mandioca.

A análise do perfil dos géis de eletroforese foi baseada na presença ou ausência da amplificação da banda, ou seja, a presença do fragmento amplificado pelos iniciadores MeWxl11-G e MeWxl11-C indicou a presença de indivíduos heterozigóticos, enquanto que a amplificação apenas de um ou outro fragmento indicou a presença de indivíduos homozigóticos para o fenótipo *waxy* (MeWxl11-G – genótipo GG) ou não *waxy* (MeWxl11-C, genótipo CC). De acordo com as análises moleculares, nenhum acesso de mandioca avaliado foi identificado na condição alélica recessiva (GG). De fato, este resultado era esperado, tendo em vista que nenhum dos 1529 acessos genotipados possui o fenótipo com amido *waxy*. Entretanto, do total de indivíduos avaliados, 206 foram identificados como heterozigóticos (CG), tendo em vista a amplificação dos fragmentos dos iniciadores MeWxl11-C e MeWxl11-G.

Avaliação genotípica e fenotípica de populações S₁ com marcadores SNAP

Vinte e oito acessos heterozigóticos foram autofecundados em campo para a geração de populações S₁ que pudessem segregar para o gene *waxy*. Deste total, três acessos (BGM0061, BGM0935 e BGM0438) produziram 187 plantas (61, 101 e 25 respectivamente) para avaliação da segregação molecular dos alelos e do fenótipo *waxy*. Os resultados da segregação alélica das progênies S₁ analisadas com os iniciadores MeWxl11-C e MeWxl11-G, demonstraram não haver desvios significativos da proporção esperada de indivíduos (1:2:1) nas progênies S₁ - BGM0061, S₁ - BGM0935 e S₁ - BGM0438, de acordo com o teste χ^2 (Tabela 4). Portanto, mesmo não havendo um grande número de indivíduos nestas três progênies, foi possível observar a segregação mendeliana dos alelos C e G localizados no íntron 11, posição 3197pb do gene *GBSSI* de acordo com o esperado para uma característica recessiva (Figura 1). Entretanto, apesar da segregação molecular ter ocorrido conforme esperado nas progênies S₁ - BGM0061, S₁ - BGM0935 e S₁ - BGM0438, a avaliação do fenótipo *waxy* nas 28 progênies S₁ em

campo, com base no teste indicativo de iodo a 2%, não identificou nenhum indivíduo com fenótipo ceroso. Portanto, este resultado evidenciou que a troca do alelo C por G na posição 3197pb do gene GBSSI, não é um bom marcador para uso na seleção assistida para identificação de indivíduos *waxy*, quando se utiliza backgrounds genéticos diferentes da fonte original da mutação AM206-5.

Tabela 4: Segregação molecular do *single nucleotide polymorphism* (SNP) na posição 3197pb (Íntron 11) do gene *Granule-bound starch synthase I – GBSSI* em três progênies S₁ de mandioca.

Família	GC	CC	GG	GC + CC		GG	χ^2
				f _o	f _e *	f _e *	
S1- BGM0061	32	16	13	48	45,75	15,25	0,44
S1- BGM0935	41	38	22	79	75,75	22,25	0,56
S1- BGM0438	10	8	7	18	18,75	6,25	0,12
Total	84	60	41				

f_o - Frequência observada e f_e – Frequência esperada de indivíduos, de acordo com a segregação mendeliana 1:2:1.

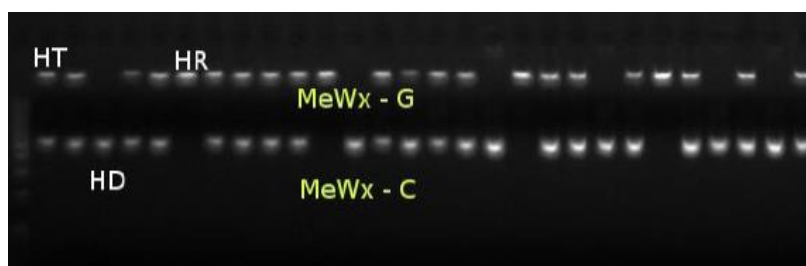


Figura 1: Exemplo de amplificação dos iniciadores MeWxI11-G e MeWxI11-C em gel de agarose 1000 2% corado com brometo de etídio na população S₁ - BGM0061 de mandioca. HT - Heterozigoto (CG) e HD – Homozigoto Dominante (CC) e HR – Homozigoto recessivo (GG).

Identificação de novas mutações no gene GBSSI por sequenciamento

Após o alinhamento das sequências e a trimagem das bases de baixa qualidade, foram identificados 16 SNPs no gene *GBSSI* e uma deleção que indicam diferenças genotípicas entre os 89 acessos de mandioca avaliados, que incluíam genótipos *waxy* e não-*waxy* (Figura 2). Deste total, seis estão localizados em regiões UTRs (*Untranslated region*), um em região não-codificante e nove em regiões codificantes. Contudo, a variação alélica identificada no gene *GBSSI* não foi capaz de identificar nenhum SNP que pudesse distinguir precisamente os

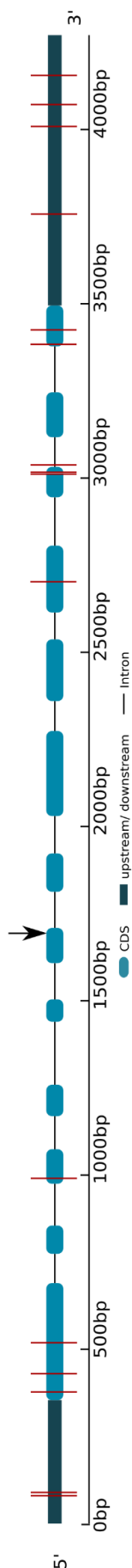


Figura 2: Representação esquemática do gene *GBSSI* (*Granule-bound starch synthase I*) indicando a presença de *single nucleotide polymorphism* (SNPs) identificados em 90 acessos de mandioca. As linhas em vermelho representam SNPs e a seta indica a deleção encontrada.

indivíduos *waxy* e não-*waxy* (Figura 3). Apenas a deleção do nucleotídeo citosina (MeWxEx6-del-C) previamente identificada por Aiemnaka et al. (2012) foi encontrada unicamente em indivíduos com fenótipo *waxy*. Portanto, os SNPs identificados no gene *GBSSI* não foram capazes de individualmente determinar códons de parada ou erros de leitura do código genético que levassem à expressão do amido *waxy* em mandioca.

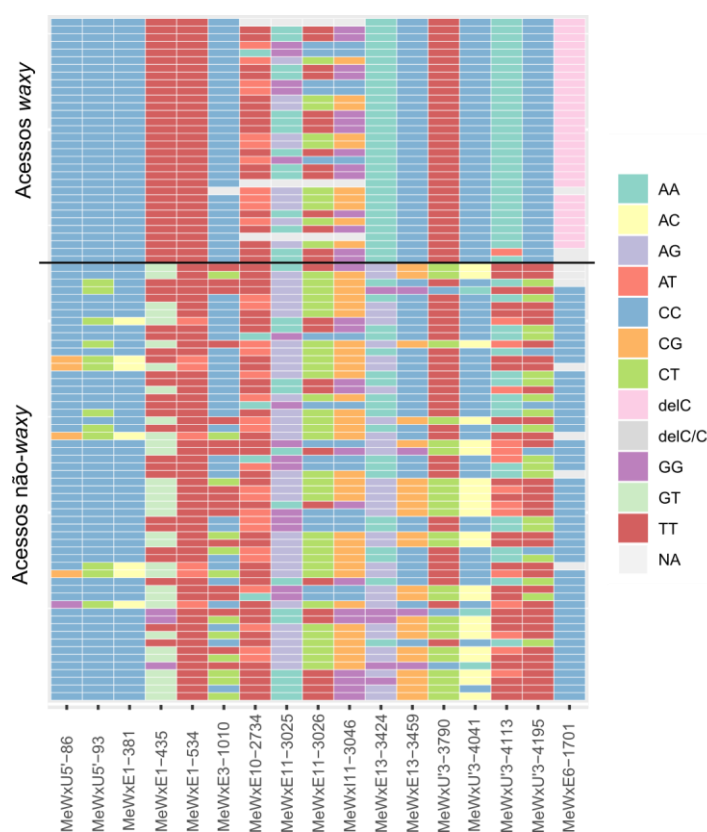


Figura 3: Variação alélica de *single nucleotide polymorphism* (SNPs) e InDel identificados no sequenciamento do gene *GBSSI* em genótipos *waxy* e não-*waxy* de mandioca. Números a frente da codificação do SNP representam sua posição no gene *GBSSI* (em pares de bases). Codificação: Me – espécie (*M. esculenta*); Wx – gene (*GBSSI*); U'5 – região UTR'5, E - regiões codificantes (éxon); I – regiões não codificantes (íntrons); U'3 – região UTR'3.

Análise discriminante de componentes principais (ADCP) dos genótipos waxy e não-waxy com base na variação alélica do gene GBSSI

Adicionalmente a ADCP foi realizada para determinar uma função discriminante para agrupamento dos diferentes indivíduos com base nos tipos de amido e no conjunto de SNPs encontrados no gene *GBSSI*. Apenas os 12 SNPs que apresentaram menor variação entre os indivíduos *waxy* e não-*waxy* foram utilizados. Os SNPs do éxon 10 (posição 2734 pb) e éxon 11 (posição 3025, 3026 e 3046 pb) foram os que mais divergiram entre os indivíduos com fenótipo *waxy* (Figura 3) e foram retirados da avaliação conjunta.

A sobreposição da densidade da função discriminante indicou que a combinação dos SNPs não foi capaz de agrupar os indivíduos *waxy* dos não-*waxy* de forma completamente acurada (Figura 4). Apesar do modelo de classificação apresentar boa acurácia não houve discriminação completa (100%) dos indivíduos nas duas classes.

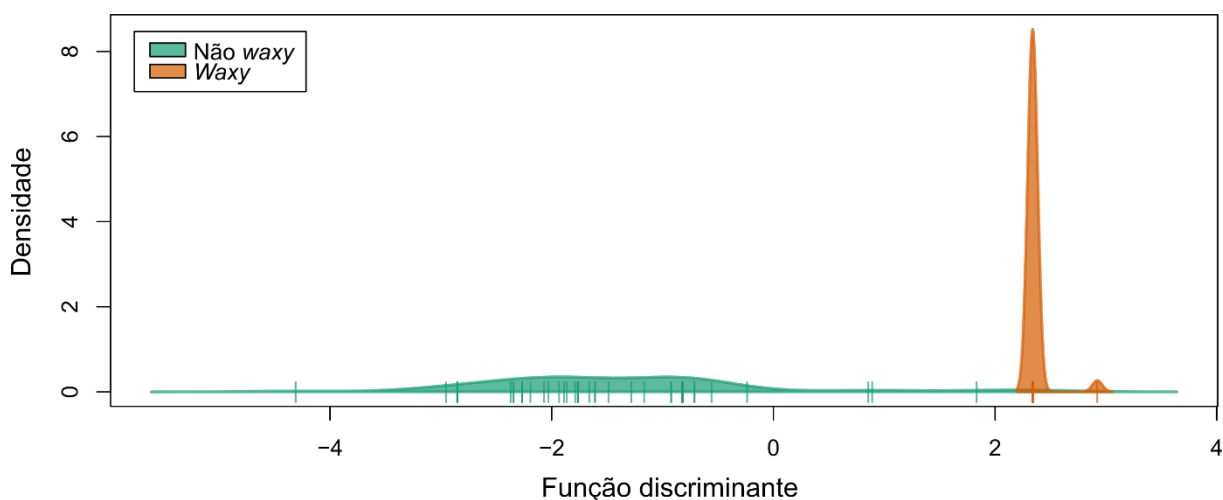


Figura 4: Gráfico de densidade da primeira função discriminante com base na análise de 12 *Single nucleotide polymorphism* (SNPs) identificados no gene *GBSSI* em indivíduos com amido *waxy* e não-*waxy*.

Avaliação genotípica do banco de germoplasma via KASP

Tendo em vista que apenas a deleção da citosina na posição 1701pb do gene *GBSSI* foi capaz de distinguir os fenótipos *waxy* dos não-*waxy* (Figura 3), o passo seguinte foi implementar um sistema de genotipagem desta deleção em larga

escala. Assim, foi realizada uma análise inicial com a técnica KASP (*Kompetitive Allele Specific PCR*) para avaliação da amplificação dos fragmentos alvo.

A fluorescência de HEX foi utilizada para detectar os acessos homozigóticos *waxy* (*wxwx*); a fluorescência FAM para detectar os acessos homozigóticos não-*waxy* (*WxWx*); enquanto que os acessos heterozigóticos foram identificados por fluorescência intermediária (*Wxwx*). A primeira etapa de treinamento da deleção MeWxEx6-del-C via KASP utilizou um conjunto de 94 acessos, e apenas os genótipos 2017wx-02-17 e 7934-1 não geraram sinal consistente para detecção pela fluorescência ou não amplificaram (Figura 5) e os genótipos 2017wx-01-02 e 2017wx-02-19 não apresentaram os alelos esperados. Portanto, esta etapa de treinamento resultou em uma taxa de sucesso de 96% na identificação dos fenótipos *waxy* e não-*waxy*. Em seguida, os 1529 acessos de germoplasma foram genotipados para validação do método e nesta etapa, apenas 18 genótipos (S1-1662-65, BGM-0052, BGM-0056, BGM-0314, BGM-0336, BGM-1037, BGM-1415, BGM-1598, BGM-1698, BGM-1716, BGM-1756, BGM-2245, BGM-2252, BGM-2257, BGM-2278, BGM-0150, BGM-0856, Brasileira) não puderam ter a condição alélica identificada (Figura 5). Assim, a KASP permitiu a identificação precisa de 98% dos indivíduos.

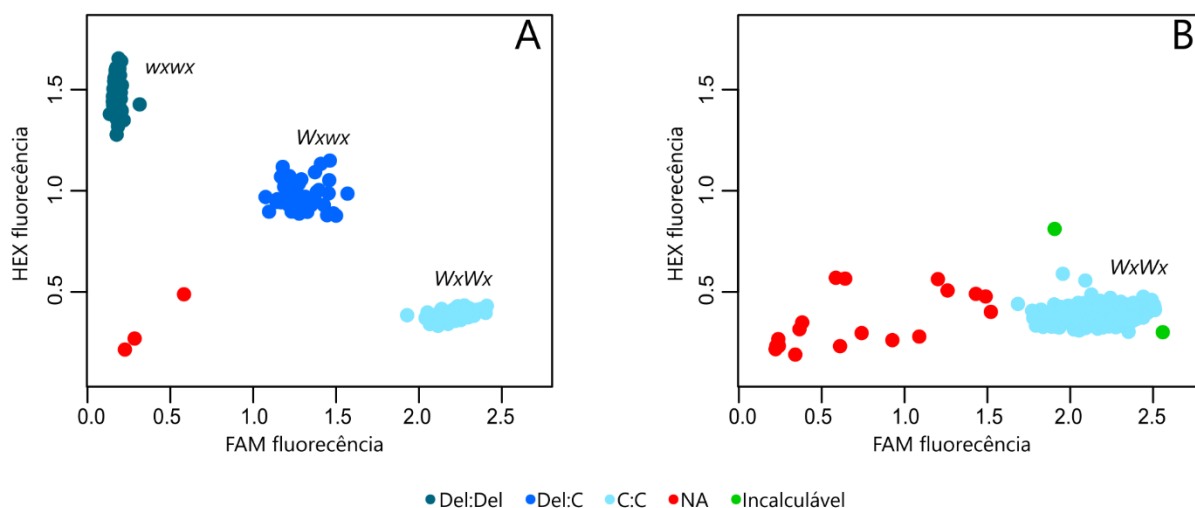


Figura 5: Genotipagem via *Kompetitive Allele Specific PCR* (KASP) para a deleção MeWxE6-del-C (A) população de treinamento composta por 94 acessos sendo 31 homozigotos amido *waxy* (*wxwx*); 31 heterozigotos amido não-*waxy* (*Wxwx*) e 32 homozigotos amido não-*waxy* (*WxWx*); e (B) população de validação composta por 1529 acessos de mandioca da América Latina.

Os principais resultados destas análises foram; i) a ausência de acessos de mandioca heterozigóticos para a deleção do alelo C na posição 1701pb do gene *GBSSI* no germoplasma avaliado; ii) a elevada especificidade na identificação da condição alélica dos genótipos na fase de treinamento sendo passível, portanto sua utilização na identificação acurada da deleção no germoplasma de mandioca da América Latina (Figura 5).

DISCUSSÃO

O principal objetivo da avaliação inicial do germoplasma de mandioca com marcadores SNAP foi identificar a presença de alelos associados ao fenótipo *waxy* e em seguida direcionar a autofecundação dos acessos uma vez que o fenótipo *waxy* é expresso sob a condição recessiva. Para isso, foram utilizados iniciadores relacionados as mutações previamente descritas por Aiemnaka et al. (2012) na fonte *waxy* AM205-6. No entanto, a ineficiência dos iniciadores MeWxl11-G/C na identificação de indivíduos com o fenótipo *waxy* no germoplasma da América Latina foi demonstrada ao avaliar as 28 famílias S₁ derivadas da autofecundação de acessos identificados como portadores do alelo G associado ao caráter. Os iniciadores reportados não identificaram o fenótipo *waxy*, uma vez que indivíduos S₁ com o genótipo G/G apresentaram alto teor de amilose, de acordo com o teste de campo do iodo. Portanto, é possível especular que o SNP previamente identificado por Aiemnaka et al. (2012) (C-G na posição 3197pb) possa ser utilizado para seleção assistida por marcadores moleculares apenas em populações segregantes derivadas da fonte AM205-6, não sendo possível validar seu uso em diferentes *backgrounds* genéticos.

A partir dos resultados para o fenótipo *waxy* pelo teste indicativo de iodo na população S₁ e segregação positiva na avaliação molecular, realizou-se o sequenciamento total do gene *GBSSI* em indivíduos *waxy* e não-*waxy*. Como consequência, todas as mutações descritas por Aiemnaka et al. (2012) (um InDel no éxon 6 – MeWxE6-del-C; duas bases no éxon 11 - MeWxE11- 3025 e 3026; e uma substituição no íntron 11 – MeWxl11 3046) foram identificados neste estudo além de outros 13 SNPs (MeWxU'5 – 86 e 93; MeWxE1 – 381,435, 543; MeWxE10 – 2734; MeWxE13 – 3424, 3459; MeWxU'3 – 3790, 4041, 4113, 4195). No entanto,

do total de 16 SNPs identificados no gene *GBSSI* nenhum deles individualmente foi capaz de distinguir os fenótipos *waxy* e não-*waxy*.

Para o sequenciamento do gene *GBSSI* foi utilizado um painel com 90 acessos de mandioca provenientes de diferentes origens da América Latina que abrigam grande diversidade genética da cultura. A diversidade dos acessos deste painel foi crucial para identificar SNPs comuns a genótipos *waxy* e não-*waxy*. Os dados são, portanto, consistentes e devem servir de base para outras investigações na cultura da mandioca em relação ao amido ceroso. Resultado semelhante também foi reportado em variedades de arroz indígenas do Nordeste da Índia onde mutações no gene *GBSSI* previamente associadas ao conteúdo de amilose (*waxy* e não-*waxy*) também não distinguiram o fenótipo ceroso (CHOUDHURY; KHAN; DAYANANDAN, 2014). Das cinco variedades de arroz *waxy* avaliadas pelos autores, duas não apresentaram a mutação, e dentre as não-*waxy* uma mutação ocorreu em uma variedade. Estes autores também investigaram o *locus* *OsC1* responsável pela coloração do apículo em arroz e encontraram inconsistência em relação as mutações e o fenótipo correspondente. Uma deleção de 10pb no gene *OsC1*, atribuída ao apículo incolor, não foi detectada em cinco das 21 variedades com este fenótipo avaliadas. Da mesma forma, uma das nove variedades com apículo colorido apresentou a deleção. Assim as mutações consideradas associadas ao amido ceroso e a coloração do apículo em arroz não corresponderam necessariamente ao fenótipo esperado em diferentes *backgrounds* genéticos.

Apenas a deleção MeWxE6-delC (posição 1701) foi exclusiva dos indivíduos cerosos que apresentaram sequenciamento nesta região derivados da fonte AM206-05 e, portanto, capaz de caracterizar o fenótipo *waxy* nestas populações. A deleção cria um códon de parada prematuro TGA no éxon 8 (posição 2069) impede a completa síntese da enzima *GBSSI* como previamente reportado por Aiemnaka et al. (2012). A enzima *GBSS* está ligada ao grânulo provavelmente sintetizando a amilose dentro da matriz granular formada pela amilopectina e mutantes com atividade reduzida ou ausente desta enzima produzem menos ou nenhuma amilose sugerindo que nenhuma outra sintase pode substituí-la nesta função (PFISTER; ZEEMAN, 2016).

A plataforma de genotipagem de SNPs alelo-específico baseada em fluorescência KASP para *screening* da deleção MeWxE6-del-C apresentou uma

acurácia de classificação dos acessos de mandioca quanto ao tipo de amido (*waxy* e não-*waxy*) de 98%. De fato, outros autores já relataram a grande flexibilidade e a elevada acurácia da técnica KASP para genotipagens de alta performance (SEMAGN et al., 2014). A genotipagem KASP permitiu clara distinção entre os acessos homozigóticos e heterozigóticos na população de treinamento, evidenciando seu grande potencial na seleção assistida por marcadores moleculares. Neste sentido, nosso trabalho se torna um importante passo para o melhoramento da cultura uma vez que valida a técnica de genotipagem KASP para detecção rotineira do fenótipo *waxy* pelo marcador MeWxEx6-del-C, quando se utiliza alelos derivados da fonte AM206-05. No futuro, será possível uma triagem de alto rendimento (*High-Throughput Screening* - HTS) com diversos marcadores funcionais importantes para a mandioca a exemplo do que vem ocorrendo em outras culturas como o trigo onde foram testados mais de 38 polimorfismos relacionados a diferentes características de importância agrônômica pela técnica KASP (RASHEED et al., 2016). A velocidade de genotipagem dos ensaios KASP foi 45 vezes superior aos marcadores PCR baseados em gel.

Análises em diferentes espécies suportam que o fenótipo *waxy* está relacionado a ausência/deficiência da enzima *GBSSI* associadas a deleções parciais e completas do gene, presença de SNPs, e elementos transponíveis, às vezes levando a códons de parada (KHARABIAN-MASOULEH et al., 2012; LARKIN; PARK, 2003; MONARI et al., 2005; SAITO; NAKAMURA, 2005; XIAOYANG et al., 2017; YI et al., 2017; ZHANG et al., 2017). Em mandioca, uma primeira mutação natural de caráter recessivo foi identificada no gene *GBSSI* caracterizando a herança deste fenótipo (AIEMNAKA et al., 2012). Todavia, esta mutação causal determinada pela deleção de uma citocina no éxon 6 na posição 1701pb não foi localizada no germoplasma analisado e, portanto, não será possível explorar o uso desta mutação para geração de populações segregantes para o gene *waxy*. Entanto, é importante destacar que ao longo do processo evolutivo diferentes pontos de mutações já foram descritos como responsáveis pelo fenótipo ceroso em outras espécies amiláceas e, portanto, outros pontos de mutações causais podem ser descobertos na cultura da mandioca, com base no sequenciamento do gene *GBSSI* em todo o germoplasma de mandioca da América Latina. Como exemplo, no milho diversas mutações relacionadas ao amido *waxy* tem disso relatadas incluindo inserções e deleções de tamanhos variados,

elementos transponíveis e SNPs (HUANG et al., 2010). As mutações mais recentes incluem a transposição do gene *rf2* (LIU; RONG; LI, 2007), deleções no éxon 7 e 10 (FAN et al., 2009) e elementos transponíveis no éxon 6 e no éxon 7 (XIAOYANG et al., 2017) identificadas como mutações funcionais no gene *GBSSI* e relacionada ao fenótipo ceroso na cultura do milho.

Outro aspecto importante é a fenotipagem de indivíduos para determinação quantitativa do conteúdo de amilose, o que poderia possibilitar a identificação de formas alélicas alternativas para diferentes teores de amilose. De acordo com Dobo et al.(2010), em arroz cinco padrões alélicos foram encontrados: TCC, TAC, GCC, GAC e GAT onde a primeira letra corresponde ao polimorfismo G/T no éxon 1, a segunda letra o polimorfismo A/C no éxon 6 e a terceira letra corresponde ao polimorfismo C/T no éxon 10. As formas alélicas TAC e TCC foram encontradas em variedades com baixo teor de amilose, GCC para as variedades com níveis intermediários de amilose aparente e para variedades com altos níveis de amilose as combinações GAT e GAC.

De modo geral, a síntese de amido é complexa e derivada da ação coordenada de várias enzimas e suas isoformas (TETLOW; MORELL; EMES, 2004). Um exemplo disso, é que recentemente uma proteína denominada *Protein Targeting To Starch* (PTST) foi identificada em *Arabidopsis thaliana* (L.) Heynh relacionada à síntese da amilose. Essa proteína atua no direcionamento e transporte da *GBSSI* (SEUNG et al., 2015, SEUNG et al., 2017) e portanto, o processo de síntese da amilose ainda não está completamente elucidado. Mutantes sem a proteína PTST em *A. thaliana* não produzem amilose no amido uma vez que a proteína *GBSS*, que normalmente se liga ao amido, não pode se ligar na ausência de *PTST*. Com isso, é possível que além das mutações diretamente em nível gênico, outras regiões genômicas possam ajudar a explicar o fenótipo *waxy* e, portanto, sejam passíveis de seleção

A identificação de diferentes mutações funcionais no gene da proteína *PTST* (Manes.02G075700) já descrito e mapeado no cromossomo 2 da mandioca, assim como na outra isoforma da *GBSSI* localizada no cromossomo 1 (Manes.01G055700) (BREDESON et al., 2016) não avaliada neste trabalho, poderá trazer novas contribuições para um melhor entendimento das bases moleculares da expressão do gene *waxy* no germoplasma da América Latina. Considerando que o gene *GBSSI* é a única sintase conhecida como responsável pela síntese de

amilose e que a proteína *PTST* é responsável pelo direcionamento da *GBSSI*, um sequenciamento completo (*paired-end*) deste genes incluindo regiões promotoras no germoplasma de mandioca seria um dos primeiros passos na busca por alelos alternativos para o fenótipo *waxy*. Toda a diversidade nucleotídica dos genes avaliados possibilitariam uma extensa avaliação *in silico* de pontos de mutações potenciais para inativação da produção de amilose. Esta informação atrelada ao resultado do sequenciamento direcionaria a autofecundação de acessos heterozigóticos. Com populações já formadas uma avaliação da segregação molecular precedida de avaliação em campo pelo teste de iodo identificaria os indivíduos promissores. Por fim, com marcadores validados em novas fontes de genótipos *waxy* uma avaliação via KASP com diferentes *backgrounds* genéticos seria realizada reafirmando o potencial destas mutações no reconhecimento de genitores potenciais de mandioca.

CONCLUSÃO

Os SNPs relacionadas ao gene *GBSSI* reportados anteriormente e os encontrados neste trabalho não são uteis para seleção assistida por marcadores moleculares em *backgrounds* genéticos diferentes da fonte AM206-5. Apenas a deleção no éxon 6 (MeWxEx6-del-C) foi capaz de discriminar completamente os genótipos *waxy* dos não-*waxy*. Apesar do germoplasma avaliado não possuir o alelo associado à fonte original AM206-5, nossos resultados contribuíram para estabelecer um modelo prático para uso da seleção assistida para variantes alélicas associadas ao fenótipo *waxy* em mandioca, de forma a otimizar a genotipagem e a identificação precoce e específica dos genótipos com os alelos desejados para geração de populações segregantes.

REFERÊNCIAS

AIEMNAKA, P.; WONGKAEW, A.; CHANTHAWORN, J.; NAGASHIMA, S. K.; BOONMA, S.; AUTHAPUN, J.; JENWEERAWAT, S.; KONGSILA, P.; KITTIPADAKUL, P.; NAKASATHIEN, S.; SREEWONGCHAI, T.; WANNARAT, W.; VICHUKIT, V.; LÓPEZ-LAVALLE, L. A. B.; CEBALLOS, H.; ROJANARIDPICHED, C.; PHUMICHA, C. Molecular characterization of a spontaneous waxy starch

mutation in cassava. **Crop Science**, v. 52, n. 5, p. 2121–2130, 2012.

BREDESON, J. V.; LYONS, J. B.; PROCHNIK, S. E.; WU, G. A.; HA, C. M.; EDSINGER-GONZALES, E.; GRIMWOOD, J.; SCHMUTZ, J.; RABBI, I. Y.; EGESI, C.; NAULUVULA, P.; LEBOT, V.; NDUNGURU, J.; MKAMILO, G.; BART, R. S.; SETTER, T. L.; GLEADOW, R. M.; KULAKOW, P.; FERGUSON, M. E.; ROUNSLEY, S.; ROKHSAR, D. S. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. **Nature biotechnology**, v. 34, n. 5, p. 562–570, 2016.

CEBALLOS, H.; SÁNCHEZ, T.; MORANTE, N.; FREGENE, M.; DUFOUR, D.; SMITH, A. M.; DENYER, K.; PÉREZ, J. C.; CALLE, F.; MESTRES, C. Discovery of an amylose-free starch mutant in cassava (*Manihot esculenta* Crantz). **Journal of Agricultural and Food Chemistry**, v. 55, n. 18, p. 7469–7476, 2007.

CHOUDHURY, B. I.; KHAN, M. L.; DAYANANDAN, S. Patterns of nucleotide diversity and phenotypes of two domestication related genes (OsC1 and Wx) in indigenous rice varieties in Northeast India. **BMC genetics**, v. 15, p. 71, 2014.

COPELAND, L.; BLAZEK, J.; SALMAN, H.; TANG, M. C. Form and functionality of starch. **Food Hydrocolloids**, v. 23, n. 6, p. 1527–1534, 2009.

DENYER, K.; JOHNSON, P.; ZEEMAN, S.; SMITH, A. M. The control of amylose synthesis. **Journal of Plant Physiology**, v. 158, n. 4, p. 479–487, 2001.

DOBO, M.; AYRES, N.; WALKER, G.; PARK, W. D. Polymorphism in the GBSS gene affects amylose content in US and European rice germplasm. **Journal of Cereal Science**, v. 52, n. 3, p. 450–456, 2010.

DOYLE, J.; DOYLE, J. Isolation of plant DNA from fresh tissue. **Focus**, v. 12, p. 13–15, 1990.

EWING, B.; GREEN, P. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. **Genome Research**, v. 8, n. 3, p. 186–194, 1998.

EWING, B.; HILLIER, L.; WENDL, M. C.; GREEN, P. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. **Genome Research**, v. 8, n. 3, p. 175–185, 1998.

FAN, L.; BAO, J.; WANG, Y.; YAO, J.; GUI, Y.; HU, W.; ZHU, J.; ZENG, M.; LI, Y.; XU, Y. Post-Domestication Selection in the Maize Starch Pathway. **PLoS ONE**, v. 4, n. 10, p. e7612, 2009.

GOODSTEIN, D. M.; SHU, S.; HOWSON, R.; NEUPANE, R.; HAYES, R. D.; FAZO, J.; MITROS, T.; DIRKS, W.; HELLSTEN, U.; PUTNAM, N.; ROKHSAR, D. S. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, n. D1, p. D1178–D1186, 2012.

HOOVER, R. Composition, molecular structure, and physicochemical properties of tuber and root starches: a review. **Carbohydrate Polymers**, v. 45, n. 3, p. 253–267, 2001.

HUANG, B. quan; TIAN, M. liang; ZHANG, J. jie; HUANG, Y. bi. waxy Locus and Its Mutant Types in Maize *Zea mays* L. **Agricultural Sciences in China**, v. 9, n. 1, p. 1–10, 2010.

JOMBART, T.; AHMED, I. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. **Bioinformatics**, v. 27, n. 21, p. 3070–3071, 2011.

KEELING, P. L.; MYERS, A. M. Biochemistry and Genetics of Starch Synthesis. **Annual Review of Food Science and Technology**, v. 1, n. 1, p. 271–303, 2010.

KHARABIAN-MASOULEH, A.; WATERS, D. L. E.; REINKE, R. F.; WARD, R.; HENRY, R. J. SNP in starch biosynthesis genes associated with nutritional and functional properties of rice. **Scientific reports**, v. 2, p. 557, 2012.

KOEHORST-VAN PUTTEN, H. J. J.; WOLTERS, A. M. a; PEREIRA-BERTRAM, I. M.; VAN DEN BERG, H. H. J.; VAN DER KROL, A. R.; VISSER, R. G. F. Cloning and characterization of a tuberous root-specific promoter from cassava (*Manihot esculenta* Crantz). **Planta**, v. 236, n. 6, p. 1955–1965, 2012.

LARKIN, P. D.; PARK, W. D. Association of waxy gene single nucleotide polymorphisms with starch characteristics in rice (*Oryza sativa* L.). **Molecular Breeding**, v. 12, n. 4, p. 335–339, 2003.

LIU, J.; RONG, T.; LI, W. Mutation loci and intragenic selection marker of the granule-bound starch synthase gene in waxy maize. **Molecular Breeding**, v. 20,

n. 2, p. 93–102, 2007.

MONARI, A. M.; SIMEONE, M. C.; URBANO, M.; MARGIOTTA, B.; LAFIANDRA, D. Molecular characterization of new waxy mutants identified in bread and durum wheat. **Theoretical and Applied Genetics**, v. 110, n. 8, p. 1481–1489, 2005.

PÉREZ, S.; BERTOFT, E. The molecular structures of starch components and their contribution to the architecture of starch granules: A comprehensive review. **Starch/Staerke**, v. 62, p. 389–420, 2010.

PFISTER, B.; ZEEMAN, S. C. Formation of starch in plant cells. **Cellular and Molecular Life Sciences**, v. 73, n. 14, p. 2781–2807, 2016.

PLUTHERO, F. G. Rapid purification of high-activity Taq DNA polymerase. **Nucleic acids research**, v. 21, n. 20, p. 4850–4851, 1993.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**, 2018. Disponível em: <<https://www.r-project.org/>>

RAEMAKERS, K.; SCHREUDER, M.; SUURS, L.; FURRER-VERHORST, H.; VINCKEN, J.-P.; DE VETTEN, N.; JACOBSEN, E.; VISSER, R. G. F. Improved Cassava Starch by Antisense Inhibition of Granule-bound Starch Synthase I. **Molecular Breeding**, v. 16, n. 2, p. 163–172, 2005.

RASHEED, A.; WEN, W.; GAO, F.; ZHAI, S.; JIN, H.; LIU, J.; GUO, Q.; ZHANG, Y.; DREISIGACKER, S.; XIA, X.; HE, Z. Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. **Theoretical and Applied Genetics**, v. 129, n. 10, p. 1843–1860, 2016.

SAITO, M.; NAKAMURA, T. Two point mutations identified in emmer wheat generate null *Wx-A1* alleles. **Theoretical and Applied Genetics**, v. 110, n. 2, p. 276–282, 2005.

SÁNCHEZ, T.; DUFOUR, D.; MORENO, I. X.; CEBALLOS, H. Comparison of pasting and gel stabilities of waxy and normal starches from potato, maize, and rice with those of a novel waxy cassava starch under thermal, chemical, and mechanical stress. **Journal of Agricultural and Food Chemistry**, v. 58, n. 8, p. 5093–5099, 2010.

ŠÁRKA, E.; DVOŘÁČEK, V. New processing and applications of waxy starch (a review). **Journal of Food Engineering**, v. 206, p. 77–87, 2017.

SEMAGN, K.; BABU, R.; HEARNE, S.; OLSEN, M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. **Molecular Breeding**, v. 33, n. 1, p. 1–14, 2014.

SEUNG, D.; BOUDET, J.; MONROE, J.; SCHREIER, T. B.; DAVID, L. C.; ABT, M.; LU, K.; ZANELLA, M.; ZEEMAN, S. C. Homologs of Protein Targeting to Starch control starch granule initiation in Arabidopsis leaves. **The Plant Cell**, v. 29, n. 7, p. 1657 LP-1677, 2017.

SEUNG, D.; SOYK, S.; COIRO, M.; MAIER, B. A.; EICKE, S.; ZEEMAN, S. C. Protein Targeting to Starch is required for localising Granule-bound starch synthase to Starch Granules and for normal amylose synthesis in Arabidopsis. **PLOS Biology**, v. 13, n. 2, p. e1002080, 2015.

SINGH, N.; SINGH, J.; KAUR, L.; SODHI, N. S.; GILL, B. S. Morphological, thermal and rheological properties of starches from different botanical sources. **Food Chemistry**, v. 81, n. 2, p. 219–231, 2003.

TETLOW, I. J.; MORELL, M. K.; EMES, M. J. Recent developments in understanding the regulation of starch metabolism in higher plants. **Journal of Experimental Botany**, v. 55, n. 406, p. 2131–2145, 2004.

UNTERGASSER, A.; CUTCUTACHE, I.; KORESSAAR, T.; YE, J.; FAIRCLOTH, B. C.; REMM, M.; ROZEN, S. G. Primer3--new capabilities and interfaces. **Nucleic acids research**, v. 40, n. 15, p. e115, 2012.

WANG, K.; HENRY, R. J.; GILBERT, R. G. Causal Relations Among Starch Biosynthesis, Structure, and Properties. **Springer Science Reviews**, v. 2, n. 1–2, p. 15–33, 2014.

WANG, S.; LI, C.; COPELAND, L.; NIU, Q.; WANG, S. Starch Retrogradation: A Comprehensive Review. **Comprehensive Reviews in Food Science and Food Safety**, v. 14, n. 5, p. 568–585, 2015.

WATERSCHOOT, J.; GOMAND, S. V.; FIERENS, E.; DELCOUR, J. A. Production, structure, physicochemical and functional properties of maize, cassava, wheat, potato and rice starches. **Starch - Stärke**, v. 67, n. 1–2, p. 14–29, 2015.

WECKX, S.; DEL-FAVERO, J.; RADEMAKERS, R.; CLAES, L.; CRUTS, M.; DE JONGHE, P.; VAN BROECKHOVEN, C.; DE RIJK, P. novoSNP, a novel computational tool for sequence variation discovery. **Genome Research**, v. 15, n. 3, p. 436–442, 2005.

XIAOYANG, W.; DAN, C.; YUQING, L.; WEIHUA, L.; XINMING, Y.; XIUQUAN, L.; JUAN, D.; LIHUI, L. Molecular characteristics of two new waxy mutations in China waxy maize. **Molecular Breeding**, v. 37, n. 3, 2017.

YI, X.; JIANG, Z.; HU, W.; ZHAO, Y.; BIE, T.; GAO, D.; LIU, D.; WU, R.; CHENG, X.; CHENG, S.; ZHANG, Y. Development of a kompetitive allele-specific PCR marker for selection of the mutated Wx-D1d allele in wheat breeding. **Plant Breeding**, v. 136, n. 4, p. 460–466, 2017.

ZEEMAN, S. C.; KOSSMANN, J.; SMITH, A. M. Starch: Its Metabolism, Evolution, and Biotechnological Modification in Plants. **Annual Review of Plant Biology**, v. 61, p. 209–34, 2010.

ZHANG, L.; CHEN, H.; LUO, M.; ZHANG, X.; DENG, M.; MA, J.; QI, P.; WANG, J.; CHEN, G.; LIU, Y.; PU, Z.; LI, W.; LAN, X.; WEI, Y.; ZHENG, Y.; JIANG, Q. Transposon insertion resulted in the silencing of Wx-B1n in Chinese wheat landraces. **Theoretical and Applied Genetics**, v. 130, n. 6, p. 1331–1331, 2017.

ZHAO, S.-S.; DUFOUR, D.; SÁNCHEZ, T.; CEBALLOS, H.; ZHANG, P. Development of waxy cassava with different Biological and physico-chemical characteristics of starches for industrial applications. **Biotechnology and Bioengineering**, v. 108, n. 8, p. 1925–1935, 2011.

ARTIGO 2

ASSOCIAÇÃO GENÔMICA AMPLA PARA IDENTIFICAÇÃO DE REGIÕES GENÔMICAS ASSOCIADAS AO AMIDO WAXY EM MANDIOCA¹

¹Artigo a ser ajustado para posterior submissão ao Comitê Editorial do periódico científico *Euphytica*, em versão na língua inglesa.

Associação genômica ampla para identificação de regiões genômicas associadas ao amido *waxy* em mandioca

Resumo: A mandioca (*Manihot esculenta* Crantz) é cultura de grande importância na produção de amido, cujos esforços recentes têm sido voltados para o melhoramento das propriedades do amido. O amido ceroso (*waxy*) de mandioca possui algumas vantagens em relação ao amido de outras fontes devido ao sabor neutro e melhores propriedades de pasta. A herança genética das fontes *waxy* conhecidas é tida como monogênica e recessiva, porém é possível que haja outras regiões genicas que possam contribuir para explicar o fenótipo *waxy*. O objetivo deste trabalho foi avaliar o potencial de uso da *Genome-Wide Association Studies* (GWAS) na detecção de variantes associadas ao amido tipo *waxy* e identificar genes ou regiões genômicas alternativos. Para tanto um painel de aproximadamente 20k de *Single Nucleotide Polymorphism* (SNP) foram utilizados numa população de 383 indivíduos (*waxy* e não-*waxy*). Todos os genótipos foram avaliados quanto à presença do amido tipo *waxy*, por meio do teste indicativo de iodo 2%. A GWAS foi conduzida pelo modelo retrospectivo *Case-control Retrospective Association Test* (CARAT) e a relação de parentesco e a estrutura da população também foram investigadas e incorporadas ao modelo. Dez SNPs foram significativos e estão localizadas no cromossomo 2 e co-localizados em regiões gênicas. Os SNPs significativos estão ainda em uma região próxima ao gene *GBSSI* e ao gene Beta-amilase. Também foram encontrados efeitos aditivos e epistáticos entre os SNPs significativos que precisam ser melhor investigados em populações segregantes de cruzamentos controlados envolvendo as fontes *waxy*. As regiões e genes identificados neste trabalho podem direcionar novos estudos e revelar processos biológicos relacionados ao fenótipo *waxy*.

Palavras-chave: *Manihot esculenta* Crantz, GBS, variável binária, interação SNP-SNP, melhoramento.

Genome-wide association studies for identification of genomic regions related to waxy starch in cassava

Abstract: Cassava (*Manihot esculenta* Crantz) is an important crop for starch production, and recently many efforts have been directed to the improvement of starch properties. The cassava waxy starch has some advantages in relation to the starch from other sources because of the neutral flavor and pasting properties. The genetic inheritance of waxy sources known is considered monogenic and recessive, however it may be possible that other genomic regions can contribute to explain the waxy phenotype. This work aims to evaluate the power of the Genome-Wide Association Studies (GWAS) to detect variants related to the waxy starch and identifying alternative genes or genomic regions. For this purpose, a panel with approximately 20k of Single Nucleotide Polymorphism (SNP) was analyzed with a set of 383 samples (waxy and non-waxy). All genotypes were evaluated in relation to the presence of waxy starch using the iodine 2% test. The GWAS was performed with the retrospective model CAse-control Retrospective Association Test (CARAT) and the kinship and population structure were also investigated and added to the model. Ten SNPs were significant, they are located in the chromosome 2 and co-located in genomic regions. The significant SNPs are in a region near to the gene GBSSI and to the gene Beta-amylase. Additive and epistatic effects were also found among the significant SNPs and need to be better investigated in segregant populations of waxy sources. The regions and genes identified in this work may drive further studies and reveal biological processes related to the waxy phenotype.

Keywords: *Manihot esculenta* Crantz, GBS, binary variable, SNP-SNP interaction, improvement.

INTRODUÇÃO

A *Manihot esculenta* Crantz é uma cultura explorada nas mais diversas condições edafoclimáticas sendo uma das mais importantes fontes de carboidratos do mundo principalmente no continente africano (RAHEEM; CHUKWUMA, 2001). Seu uso comercial se concentra na produção de raízes para a produção de amido conhecido popularmente no Brasil como fécula (CEBALLOS et al., 2004). O amido é basicamente constituído de dois tipos de polímeros: a amilose (essencialmente α 1,4-poliglucanos - lineares) e a amilopectina (α 1,4-poliglucanos e α -1,6-poliglucanos - ramificados) cuja proporções determinam suas características e funcionalidades (PÉREZ; BERTOFT, 2010). Os extremos desta relação são os mais explorados e amidos com baixo teor de amilose em mandioca (amido tipo ceroso - *waxy*) são preferíveis principalmente na indústria de alimentos devido a sua menor retrogradação, baixo teor de lipídios, alta viscosidade, além da produção de géis claros com fácil gelatinização (MOORTHY, 2002; SÁNCHEZ et al., 2010)

A herança genética das fontes atuais que possuem o fenótipo *waxy* é monogênica recessiva (CEBALLOS et al., 2007) e uma deleção natural cria um códon de parada no gene *GBSSI* (*Granule-bound starch synthase I*) (AIEMNAKA et al., 2012) caracterizando o fenótipo *waxy* na mandioca. Além disso, o silenciamento do gene *GBSSI* também foi capaz de produzir mutantes transgênicos com o fenótipo *waxy* (ZHAO et al., 2011). Em se tratando de uma herança monogênica, a transferência do gene *waxy* para variedades elite de mandioca seria uma atividade relativamente simples. Contudo, a alta heterozigose inerente da espécie faz com que a transferência de características monogênicas, como é o caso do gene *waxy*, via retrocruzamentos seja bastante complexa. Isto porque geralmente ocorre uma alta recombinação genética não sendo possível recuperar por completo o genoma do parental recorrente (CEBALLOS; KULAKOW; HERSHEY, 2012). Desta forma, esforços para o melhoramento desta característica estão concentrados em duas frentes: *i*) desenvolver metodologias rápidas que possibilitem a seleção indireta e precoce da característica permitindo uma redução no ciclo de melhoramento da mandioca; e *ii*) identificar variantes potenciais para novos pontos de mutação e diversificar as fontes de amido *waxy*.

Apesar das enzimas envolvidas na biossíntese de amido serem bem conservadas sugerindo uma origem comum (DESCHAMPS et al., 2008), diferentes mecanismos operaram no gene *waxy* durante a evolução (MONARI et al., 2005),

tornando complexa a compreensão do fenótipo entre as diferentes culturas amiláceas. A relação entre a variação natural em genes relacionados a síntese de amido e suas propriedades tem sido melhor estabelecidas em outras espécies. Elementos transponíveis, inserções e deleções de diferentes comprimentos e SNPs (*Single Nucleotide Polymorphism*) são correlacionados ao fenótipo *waxy* em diferentes espécies como arroz (KHARABIAN-MASOULEH et al., 2012; LARKIN; PARK, 2003), milho (XIAOYANG et al., 2017) e trigo (MONARI et al., 2005; SAITO; NAKAMURA, 2005; ZHANG et al., 2017). De maneira complementar, em culturas como o trigo e arroz o efeito de dose de cada umas das formas alélicas já foi caracterizado entre mutantes *waxy* parciais e completos (DOBO et al., 2010; YAMAMORI; QUYNH, 2000).

Adicionalmente, efeitos pleiotrópicos observados em mutantes de enzimas envolvidas no metabolismo do amido indicam que estas podem operar dentro de complexos proteicos (FUJITA et al., 2007; GEERA et al., 2006; TETLOW; MORELL; EMES, 2004). Nos últimos anos novas proteínas tem sido identificadas envolvidas na estrutura e/ou quantidade dos constituintes do amido (PFISTER; ZEEMAN, 2016; XING et al., 2016). Por exemplo, a proteína plastidial *protein targeting to starch* (PTST) foi reportada como estando envolvida no processo de síntese da amilose em *Arabidopsis* por direcionar a enzima *GBSSI* para os locais de síntese da amilose (SEUNG et al., 2015, 2017). Mutantes *ptst* sintetizam amido sem amilose sendo fenotipicamente semelhantes aos mutantes que não possuem *GBSSI* e portanto, podem ser novos candidatos para explicar este fenótipo em mandioca.

Assim, genes alternativos ou regiões genômicas diferentes dos relatados anteriormente, precisam ser melhores investigadas para verificar novas associações com o fenótipo *waxy*. Neste sentido, com o avanço das novas plataformas *Next Generation Sequencing* (NGS) e a genotipagem de alta densidade de SNPs, o custo destas análises foi bastante reduzido tornando-se possível a utilização de grandes painéis de marcadores para identificação de associações entre marcadores e características agronomicamente importantes. Estudos *GWAS* em mandioca são recentes e basicamente focados para resistência ao vírus do mosaico africano (WOLFE et al., 2016), a podridão radicular (BRITO; OLIVEIRA; OLIVEIRA, 2017), ao vírus do castanho listrado da mandioca (KAYONDO et al., 2018), bem como para características de raízes como conteúdo de carotenoides (ESUMA et al., 2016) e de matéria-seca nas raízes (RABBI et al., 2017). Desta forma, o objetivo deste trabalho foi realizar a Genotipagem por Sequenciamento (GBS) em acessos de mandioca para avaliar o potencial de uso da *GWAS* na detecção de variantes associadas ao amido

tipo *waxy*. Adicionalmente foi investigada a relação dos genes co-localizados com a rota metabólica do amido e a interação SNP-SNP entre os marcadores com associação ao amido *waxy*.

MATERIAL E MÉTODOS

Material vegetal e fenotipagem dos genótipos

Foram analisados 352 acessos pertencentes ao Banco de germoplasma de mandioca da Embrapa Mandioca e Fruticultura (CNPMPF), juntamente com 31 híbridos com amido tipo *waxy* provenientes de cruzamentos com a fonte original AM206-5 (CEBALLOS et al., 2007) (Material Suplementar, Tabela S1).

Todos os genótipos foram confirmados quanto à presença ou não do amido tipo *waxy*, por meio do teste indicativo de iodo a 2% (2g KI e 0.2g I₂ em água destilada), aplicado em uma secção transversal das raízes de todos os genótipos durante o período de colheita, que variou de 10 a 12 meses (Material Suplementar, Tabela S1). Como resultado do teste a coloração azul-escura indica a presença de amilose e a coloração marrom-avermelhada indica ausência/baixo teor de amilose (DENYER et al., 2001).

Genotipagem por sequenciamento, controle de qualidade e imputação

A extração de DNA genômico foi realizada segundo protocolo CTAB (brometo de cetiltrimetilamônio) (DOYLE; DOYLE, 1990). As amostras foram padronizadas a 60 ng/μL e enviadas para *Cornell Biotechnology Resource Center* para genotipagem por sequenciamento – GBS (ELSHIRE et al., 2011) utilizando a enzima *ApeKI*. A GBS foi realizada utilizando a plataforma *Genome Analyzer 2000* (Illumina, Estados Unidos) e o alinhamento das leituras realizado com o genoma de referência da mandioca v6.1 (BREDESON et al., 2016) depositado no banco de dados do Phytozome v11.0 com auxílio do programa TASSEL v.5 (BRADBURY et al., 2007).

Todos os indivíduos passaram pelo filtro de *call rate* < 0,4 e apenas SNPs com *call rate* de <0,1 permaneceram na análise. A imputação foi implementada pelo programa BEAGLE v.5 que utiliza o modelo oculto de Markov (HMM) para inferir agrupamentos de haplótipos e ainda proceder a imputação dos dados faltantes (BROWNING; BROWNING, 2013, 2016). Após a imputação, SNPs com menor frequência alélica (MAF) > 0,01 foram excluídos da análise totalizando 20.956 SNPs

e 383 genótipos. Ambos filtros de qualidade foram executados com auxílio do programa PLINK v.1.07 (PURCELL et al., 2007).

Desequilíbrio de ligação

A extensão do desequilíbrio de ligação (DL) para todos os 18 cromossomos foi estimada utilizando os coeficientes de correlação (r^2) entre os pares de loci com auxílio do pacote LDheatmap (SHIN et al., 2006) no ambiente de programação R v3.5 (R CORE TEAM, 2018). Para investigar o declínio do DL, os valores de r^2 foram plotados em função da distância física, em pares de base (pb) e em seguida, realizada uma regressão não linear (WEISBERG, 2005) implementada pela função “nls” no ambiente de programação R v3.5 (R CORE TEAM, 2018).

Análise de estrutura populacional e parentesco genômico

A estrutura populacional foi inferida pelo algoritmo bayesiano variacional fastStructure (RAJ; STEPHENS; PRITCHARD, 2014) com diferentes números de K ($K=2-10$). Para escolher o número apropriado de populações (complexidade do modelo) o programa utiliza: (1) o valor de K que maximiza o limite inferior de probabilidade log-marginal - LLBO de todo o conjunto de dados; e (2) o valor limite, do menor número de componentes do modelo que representa quase todos os ancestrais da amostra. Tais métricas fornecem uma faixa de valores para o número de populações que explicam a estrutura dos dados. Assim, a matriz Q foi obtida com base na menor complexidade do modelo na qual os erros de previsão não variam significativamente (RAJ; STEPHENS; PRITCHARD, 2014).

A matriz de parentesco genômica - *Kinship* (K) foi calculada com base no método de Vanraden (2008), na qual $U = \frac{ZZ'}{2\sum p_i(1-p_i)}$, em que $Z = M - P$, sendo M a matriz dos marcadores e P a matriz das frequências alélicas expressas por $2(p_i - 0,5)$ e agrupamento hierárquico foi realizado pelo método de ligação completa implementados com o auxílio do pacote GAPIT (LIPKA et al., 2012; TANG et al., 2016) no ambiente de programação R v3.5 (R CORE TEAM, 2018).

Análise de associação genômica ampla

A GWAS foi conduzida pelo modelo retrospectivo CARAT v1.2 (*Case-control Retrospective Association Test*) onde a distribuição de probabilidade do genótipo e

das covariáveis é construída a partir do fenótipo. O CARAT é baseado em uma estrutura de quase-probabilidade de efeitos mistos onde o fenótipo é a variável independente. Para capturar a estrutura da população, a matriz de parentesco genômica foi incorporada como um efeito poligênico aditivo e a estrutura populacional inferida pelo fastStructure foi adicionada ao modelo como covariável.

Associações significativas entre os SNPs e os fenótipos foram estimadas pelo teste *Bonferroni* ($p < 0,01$). Gráficos Quantile-Quantile e Manhattan foram elaborados com auxílio do pacote “qqman” (TURNER, 2018) implementado no ambiente de programação R v3.5 (R CORE TEAM, 2018).

A interação entre os SNPs significativos foi implementada com o auxílio do pacote *SNPassoc* (GONZÁLEZ et al., 2007, 2014) no ambiente de programação R v3.5 (R CORE TEAM, 2018). As interações foram avaliadas pelo teste da razão log-verossimilhança (LTR). Para cada par de SNPs (i, j), foi calculada a probabilidade subjacente ao modelo nulo L_0 ; a probabilidade sob cada *single-SNP*, $L_{(i)}$ e $L_{(j)}$; a probabilidade sob um o modelo SNP aditivo $L_{a(i,j)}$; e a probabilidade sob um modelo SNP completo (incluindo interação a SNP-SNP) $L_{f(i,j)}$. A partir dessas probabilidades, a interação epistática foi obtida por $LTR_{i,j} = -2(\log L_{f(i,j)} - \log L_{a(i,j)})$; o efeito bruto de cada SNP por $LTR_{ii} = -2(\log L_i - \log L_0)$; e a probabilidade aditiva de dois SNPs em relação ao melhor dos modelos *single-SNP* por $LTR_{ji} = -2(\log L_{a(i,j)} - \log \max(L_i L_j))$ (GONZÁLEZ et al., 2007, 2014). Para todos os testes de interações entre SNPs, p -valor $< 0,01$, foram considerados significativos.

Avaliação in silico

SNPs significativos foram localizados no genoma de referência da mandioca v6.1 (BREDESON et al., 2016) depositado no banco de dados do Phytozome v11.0 pela ferramenta JBrowse (GOODSTEIN et al., 2012). Genes referentes a rota metabólica do amido relacionados a sua síntese (Mesculenta PWY-6724) e degradação (Mesculenta PWY-622) foram utilizados para visualizar a proximidade com os polimorfismos significativos encontrados neste trabalho (Material suplementar, Tabela S2). A anotação funcional de todos os genes foi identificada pela ferramenta PhytoMine do Phytozome v11.0 (GOODSTEIN et al., 2012).

RESULTADOS

Desequilíbrio de ligação (LD) e cobertura genômica da genotipagem

O padrão de decaimento do LD sobre a distância física determina a densidade de cobertura necessária para análises genômicas como a GWAS (FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003). O LD médio obtido a partir dos 12.857.123 pares de SNPs foi de $r^2=0,021$. O r^2 médio variou bastante entre os diferentes cromossomos de 0,029 no cromossomo 18 a 0,016 no cromossomo 10. Apenas 1,3% dos pares de SNPs, apresentaram $r^2 > 0,2$, enquanto que 0,4% e 0,05% apresentam r^2 superiores a 0,5 e 0,9, respectivamente. Como esperado houve um rápido decaimento do LD, com uma distância física de aproximadamente 100pb considerando um $r^2 = 0,2$ (Figura 1) uma vez que as amostras amido *waxy* são provenientes de autofecundação.

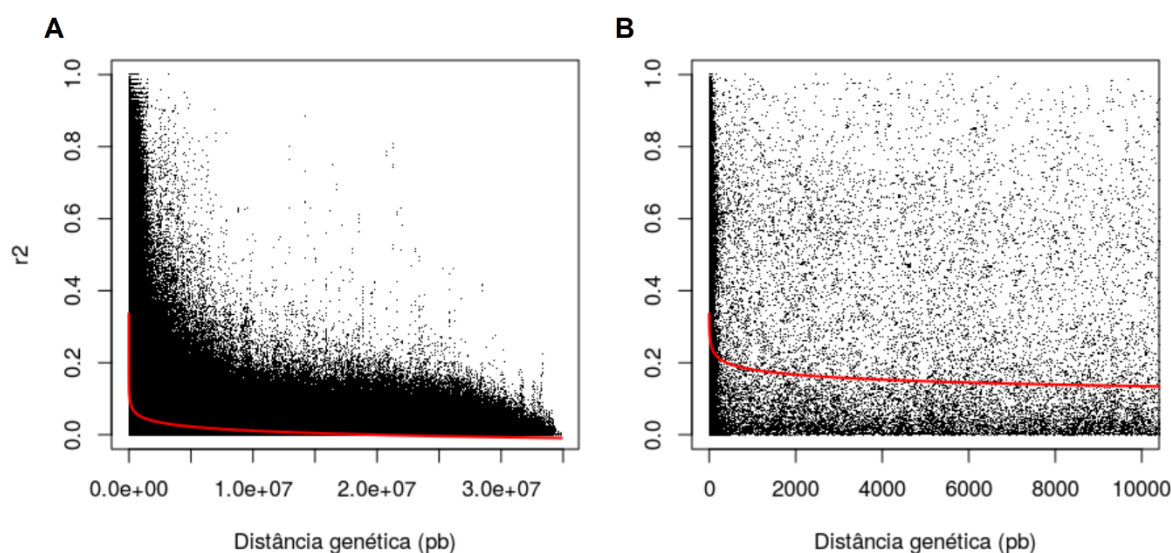


Figura 1: Padrão de decaimento do desequilíbrio de ligação em relação à distância entre marcadores *Single Nucleotide Polymorphism* (SNP), considerando a análise conjunta dos 18 cromossomos da mandioca, identificados em um painel com 383 genótipos. Os pontos em preto indicam os valores de r^2 de todos os pares de SNPs. A curva vermelha representa uma função não linear de r^2 contra a distância física do SNP (Mb). (A) Visão geral do decaimento; (B) Aproximação com base na distância até 10kb.

Apesar do rápido decaimento do LD as principais regiões genômicas onde os genes relacionados a rota metabólica do amido estão localizados apresentou ampla cobertura (Figura 2). A densidade média dos polimorfismos foi de 24kpb em todo o

genoma com destaque para o cromossomo um com a maior quantidade de SNPs (2060) e maior cobertura (1 SNP a cada 16kpb) e o cromossomo 17 com menor quantidade de SNPs (773) e cobertura (1 SNP a cada 35kpb).

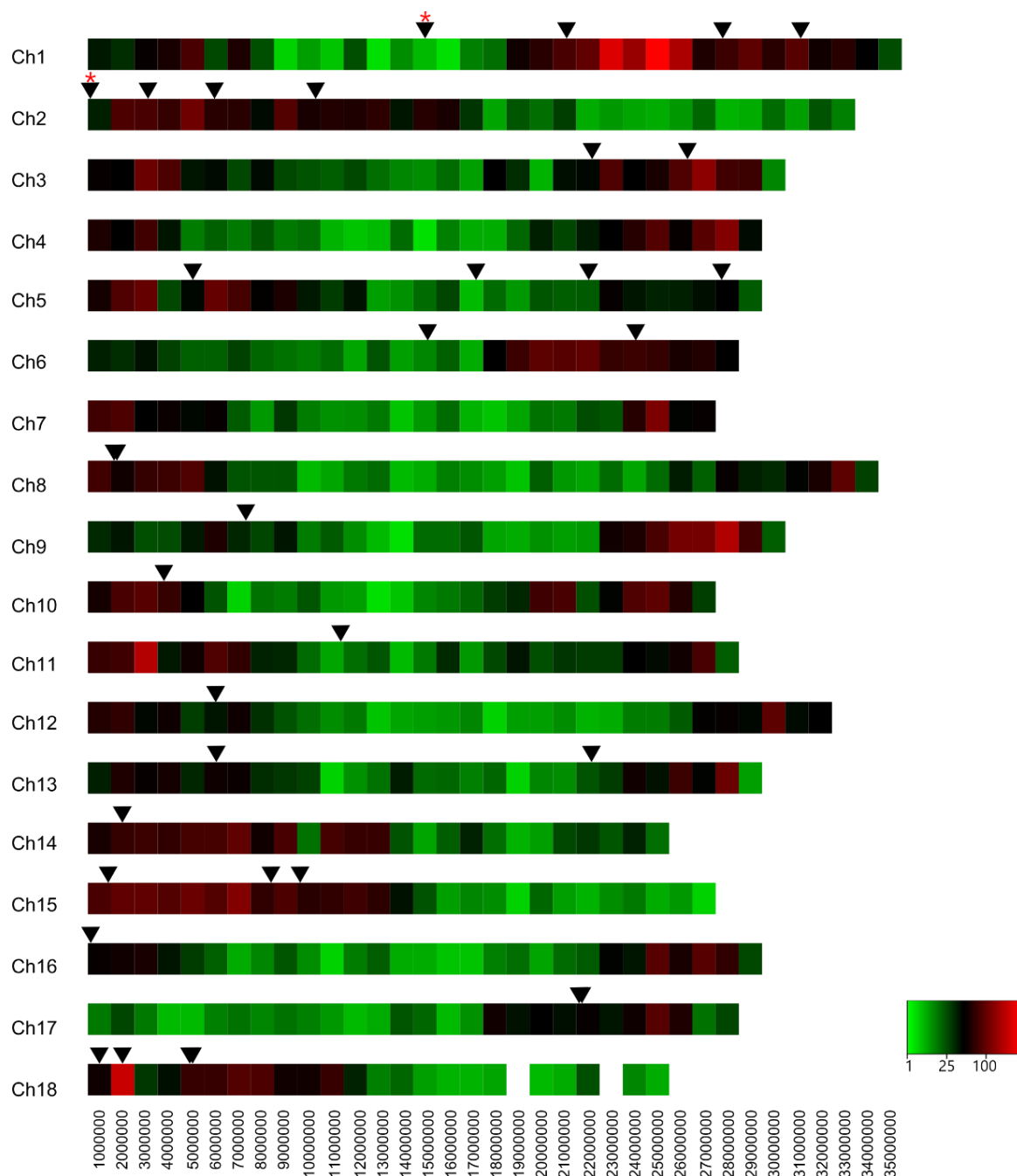


Figura 2: Distribuições de *Single Nucleotide Polymorphism* (SNP) nos 18 cromossomos da mandioca com a co-localização de genes relacionados a rota metabólica do amido (setas). O eixo horizontal exibe o comprimento do cromossomo em pares de bases; Legenda indica a densidade de SNPs. Asteriscos indicam as posições do gene *Granule-bound starch synthase I* – GBSSI.

No entanto, houve uma baixa densidade de SNPs flanqueando os genes: GBSSI no cromossomo 1 (Manes.01G055700); enzima de ramificação de glicogênio no cromossomo 5 (Manes.05G133800); e uma ADP-glicose sintase no cromossomo 11 (Manes.11G085500). Também houveram *gaps* no cromossomo 18 com duas áreas de 100k sem nenhum SNP polimórfico. Porém, não existem genes relacionados à rota metabólica do amido relatados nesta região (Figura 2). Portanto, as regiões de maior interesse para as avaliações sobre a rota metabólica do amido estão com cobertura genômica adequada.

Estrutura da população e parentesco

A matriz de parentesco genômico demonstrou a existência de três grandes grupos de genótipos com presença de subdivisões da população (Figura 3). O grupo um é formado por indivíduos com alta similaridade genética e mais distantes em relação aos demais grupos. O grupo dois com 180 genótipos possui pelo menos cinco subdivisões evidentes. O terceiro agrupamento, no entanto, apresentou em sua maioria genótipos com baixa similaridade genética, à exceção dos genótipos *waxy*, certamente pelo fato de serem clones derivados de cruzamentos com certo grau de parentesco. A estratificação populacional inferida pelo algoritmo do fastStructure indicou a presença de quatro grupos explicando a complexidade do modelo convergindo com a matriz de parentesco. Entretanto, é interessante destacar que os indivíduos com o fenótipo *waxy* na matriz de estrutura compartilharam alelos com todos os demais grupos (Figura 3).

Associação genômica ampla

Foram detectadas dez associações estatisticamente significativas após a correção de Bonferroni ($p < 0,01$), entre marcadores SNP e o fenótipo *waxy* (Figura 4). Todos os SNPs significativos foram alocados em regiões gênicas localizados no cromossomo dois (Tabela 1). Apesar da matriz de estrutura da população ter sido incluída no modelo, o QQ-plot ainda permaneceu inflado para alguns SNPs indicando que muitos p-valores moderadamente significativos são mais significativos do que o esperado sob a hipótese nula (Figura 4).

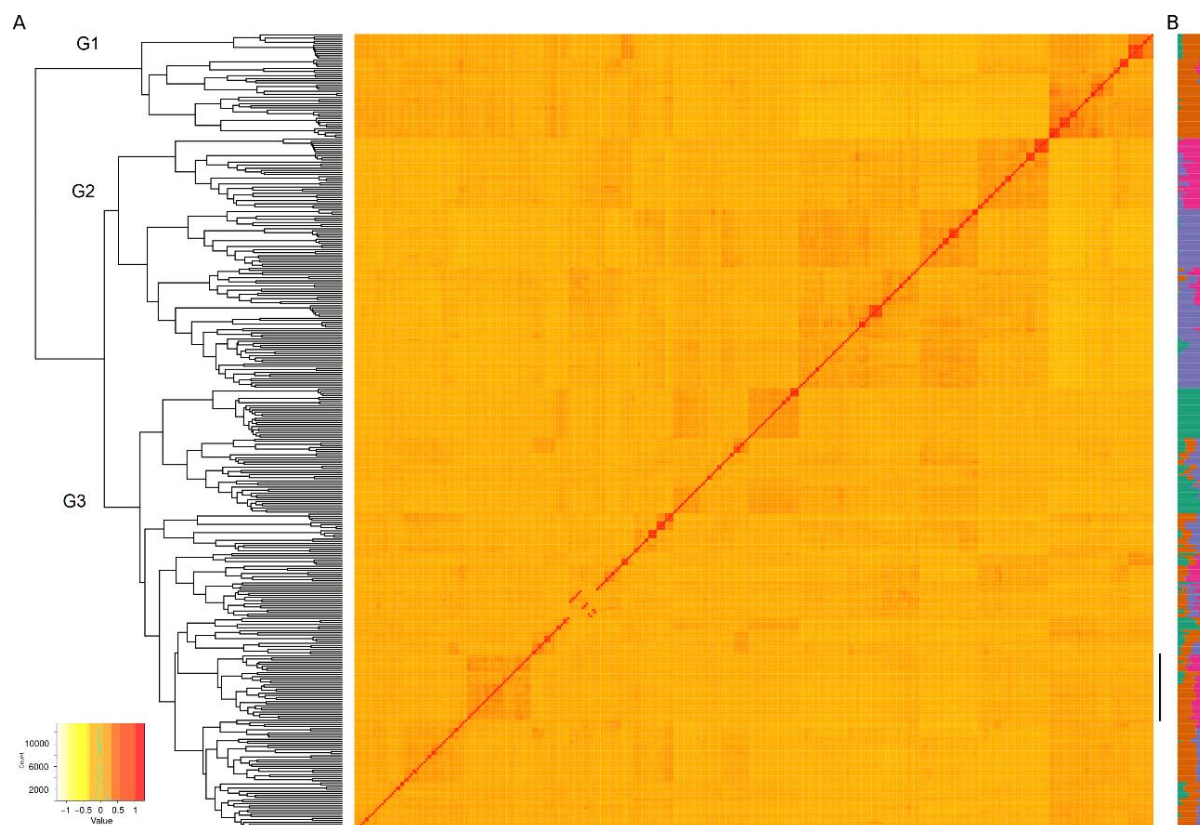


Figura 3: Estrutura populacional com base na análise de 20,956 *Single Nucleotide Polymorphism* (SNP). (A) relação entre acessos de mandioca baseada em matriz de parentesco genômico, e (B) agrupamento inferido por análise bayesiana para $k=4$. Os genótipos são mostrados na linha vertical e cada grupo genético tem uma cor distinta. Cada linha vertical representa um indivíduo dividido em segmentos cuja cor e tamanho são proporcionais à porcentagem de alelos compartilhados em cada grupo. Em destaque genótipos com amido tipo *waxy*.

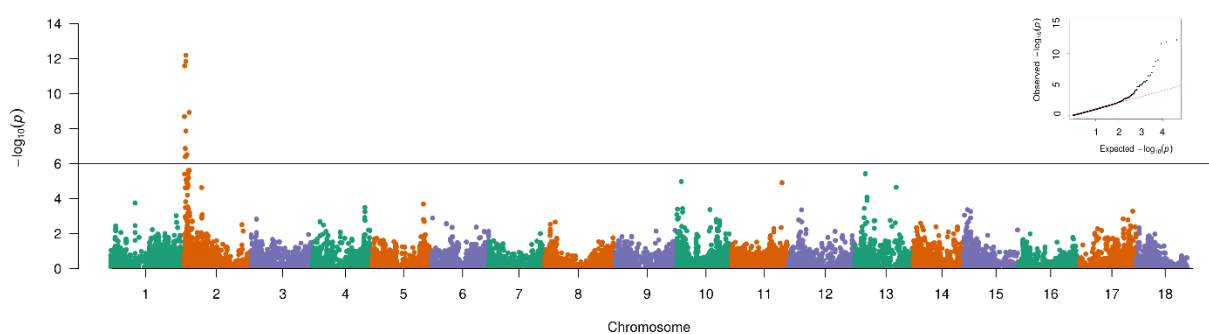


Figura 4: Gráfico *Manhattan plot* indicando *Single Nucleotide Polymorphism* (SNP) associados ao fenótipo *waxy*. O eixo y representa o p -valor da associação marcador-característica em uma escala $-\log_{10}$. A linha indica o valor do *threshold* para declarar associação significativa, de acordo com a correção de Bonferroni ($p < 0,01$). Gráfico Quantile-Quantile (QQ) dos p -valores observados da análise de associação e esperados numa associação nula para o fenótipo.

Tabela 1: Relação de marcadores *single-nucleotide polymorphism* (SNP) com associação marcador-característica corrigidos pelo teste de Bonferroni ($p < 0,01$) para o fenótipo *waxy* em genótipos de mandioca, localizados no cromossomo 2 em ordem de significância.

SNP	<i>p</i> -valor (-log10)	<i>Waxy</i> (%)			Não <i>waxy</i> (%)			Transcrito	Função
		<i>A</i> ₁ <i>A</i> ₁	<i>A</i> ₂ <i>A</i> ₂	<i>A</i> ₁ <i>A</i> ₂	<i>A</i> ₁ <i>A</i> ₁	<i>A</i> ₂ <i>A</i> ₂	<i>A</i> ₁ <i>A</i> ₂		
S2_1148905	12.1957	3.23	93.55	3.23	94.32	0.00	5.68	Manes.02G013600	Fator de iniciação de tradução
S2_1149831	11.8457	3.23	93.55	3.23	94.89	1.14	3.98	Manes.02G013600	Fator de iniciação de tradução
S2_592034	11.5917	3.23	93.55	3.23	94.03	0.28	5.68	Manes.02G005700	<i>ATP-binding</i>
S2_2783349	8.9342	6.45	70.97	22.58	92.61	1.14	6.25	Manes.02G035900	Atividade hidrolase
S2_467229	8.6923	3.23	87.10	9.68	92.05	0.57	7.39	Manes.02G003500	Função desconhecida
S2_1149750	7.8647	3.23	93.55	3.23	84.38	5.11	10.51	Manes.02G013600	Fator de iniciação de tradução
S2_892913	6.8703	3.23	93.55	3.23	78.69	3.41	17.90	Manes.02G009400	Função desconhecida
S2_1660084	6.5136	0.00	87.10	12.90	75.57	3.13	21.31	Manes.02G020800	Função desconhecida
S2_894536	6.3976	3.23	87.10	9.68	80.11	5.40	14.49	Manes.02G009400	Função desconhecida
S2_894562	6.3976	3.23	87.10	9.68	80.11	5.40	14.49	Manes.02G009400	Função desconhecida

Genes co-localizados com SNPs significativos para o fenótipo waxy

Todos os SNPs significativos pela GWAS foram co-localizados em regiões gênicas. No entanto, os transcritos Manes.02G003500, Manes.02G009400 e Manes.02G020800 não possuem função proteica descrita até o presente momento. Dos dez SNPs significativos, três estão localizados no gene responsável por um fator de iniciação de tradução IF-2 mitocondrial (Manes.02G013600), um SNP com atividade de ligação de ATP ABCF3 (Manes.02G005700) e outro com atividade hidrolase (Manes.02G035900) (Figura 5A). Os SNPs significativos S2_1148905, S2_1149831 e S2_1149750 relacionados ao gene Manes.02G013600 possuem alto desequilíbrio de ligação e o SNP S2_1660084 de função desconhecida está próximo a um bloco haplótipo (Figura 5B).

Em relação a genes próximos aos SNPs significativos dois deles referem-se a rota metabólica do amido: GBSSI (Manes.02G001000) e Beta-amilase (Manes.02G006200) (Figura 5A). A enzima GBSSI (EC 2.4.1.242) está inserida na rota de síntese de amido (*Mesocultura* PWY-6724) e é diretamente responsável pela síntese da amilose. Foi descrita primeiramente por Aiemnaka et al. (2012) no cromossomo dois e uma outra isoforma está localizada no cromossomo 1 (Manes.01G055700) descrita após novo sequenciamento da espécie (BREDESON et al., 2016). Em contrapartida a enzima Beta-amilase (EC 3.2.1.2) está inserida na rota de degradação do amido de mandioca (*Mesocultura* PWY-622) e atua no amido produzindo beta-maltose por meio de uma inversão. A enzima beta-amilase possui isoformas no cromossomo três (Manes.03G155800), cinco (Manes.05G034800, Manes.05G034900, Manes.05G073400, Manes.05G189000, Manes.05G194300), seis (Manes.06G054000), oito (Manes.08G091100, Manes.08G162200), doze (Manes.12G078500, Manes.12G142500), treze (Manes.13G026800), quatorze (Manes.14G122700), e no cromossomo quinze (Manes.15G047700, Manes.15G171200, Manes.15G190300) na cultura da mandioca.

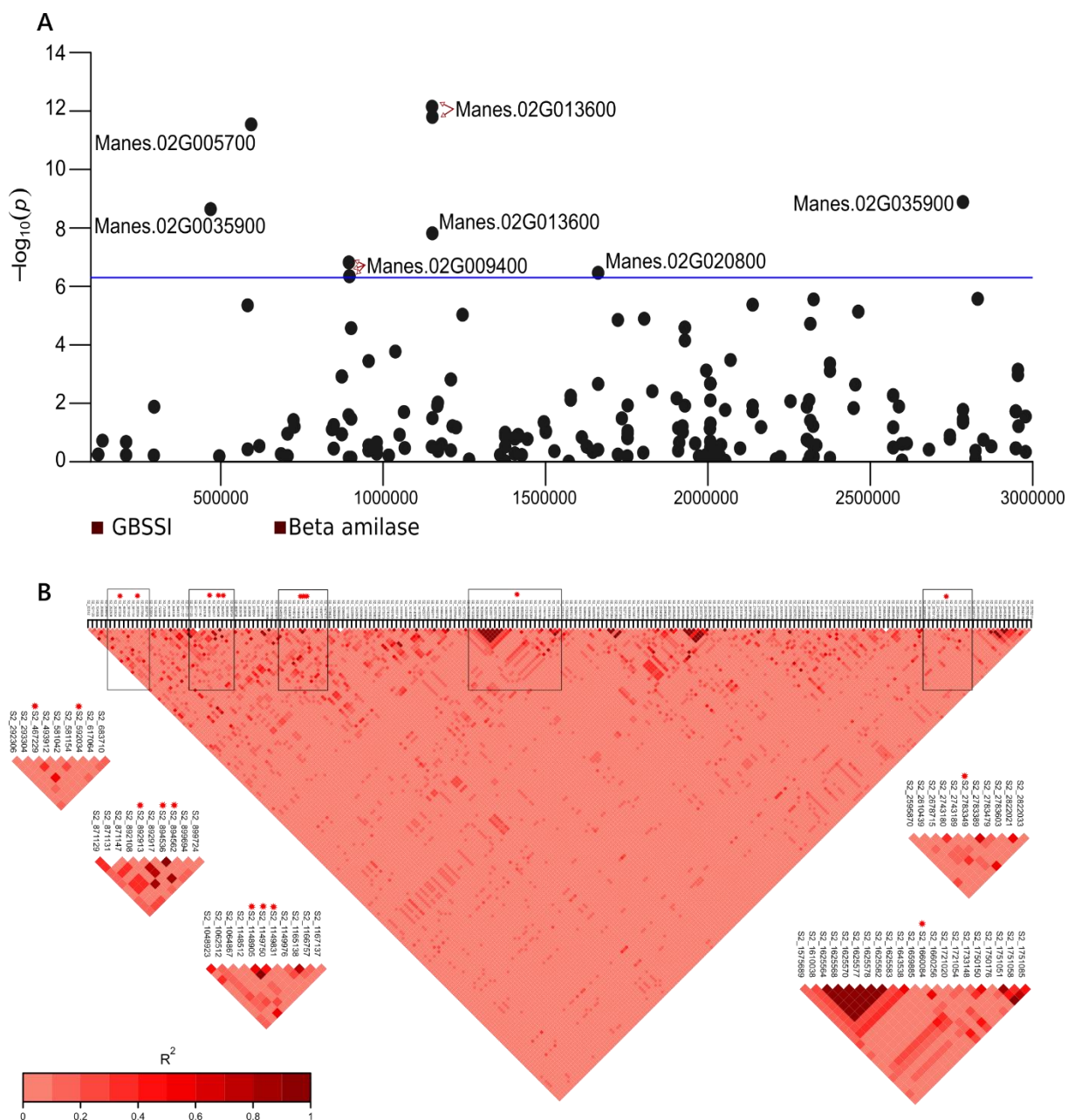


Figura 5: Associação genômica do cromossomo 2. A - Indicação dos genes colocalizados aos *Single Nucleotide Polymorphism* (SNP) significativos. O eixo y representa o p -valor da associação marcador-fenótipo em uma escala $-\log_{10}$. A linha indica o valor do *threshold* para declarar associação significativa, de acordo a correção de Bonferroni ($p < 0,01$). Caixa vermelha indica posição dos genes referentes a rota metabólica do amido (pb). B - Desequilíbrio de ligação com valores de r^2 de todos os pares de SNPs correspondente à faixa de SNPs no cromossomo dois. Em destaque desequilíbrio de ligação dos SNPs significativos pela GWAS indicados por asterisco.

Interação SNP-SNP

Para avaliar a interação entre os SNPs significativos e o fenótipo foi realizada uma análise de interação sob o modelo recessivo pelo teste de razão log-verossimilhança (Figura 6). A maioria dos SNPs não apresentou efeito epistático significativo a exceção dos pares S2_1660084 x S2_892913, S2_114950 x S2_892913 e S2_1149831 x S2_1149750 ($p < 0,01$). Em contrapartida, o efeito de aditividade foi altamente significativo ($p < 0,01$) em 24 das 45 combinações de SNPs. Os SNPs S2_1149750 e S2_273349 foram altamente significativos em sete das nove combinações, enquanto que o SNP S2_1148905 não apresentou significância com os demais polimorfismos avaliados.

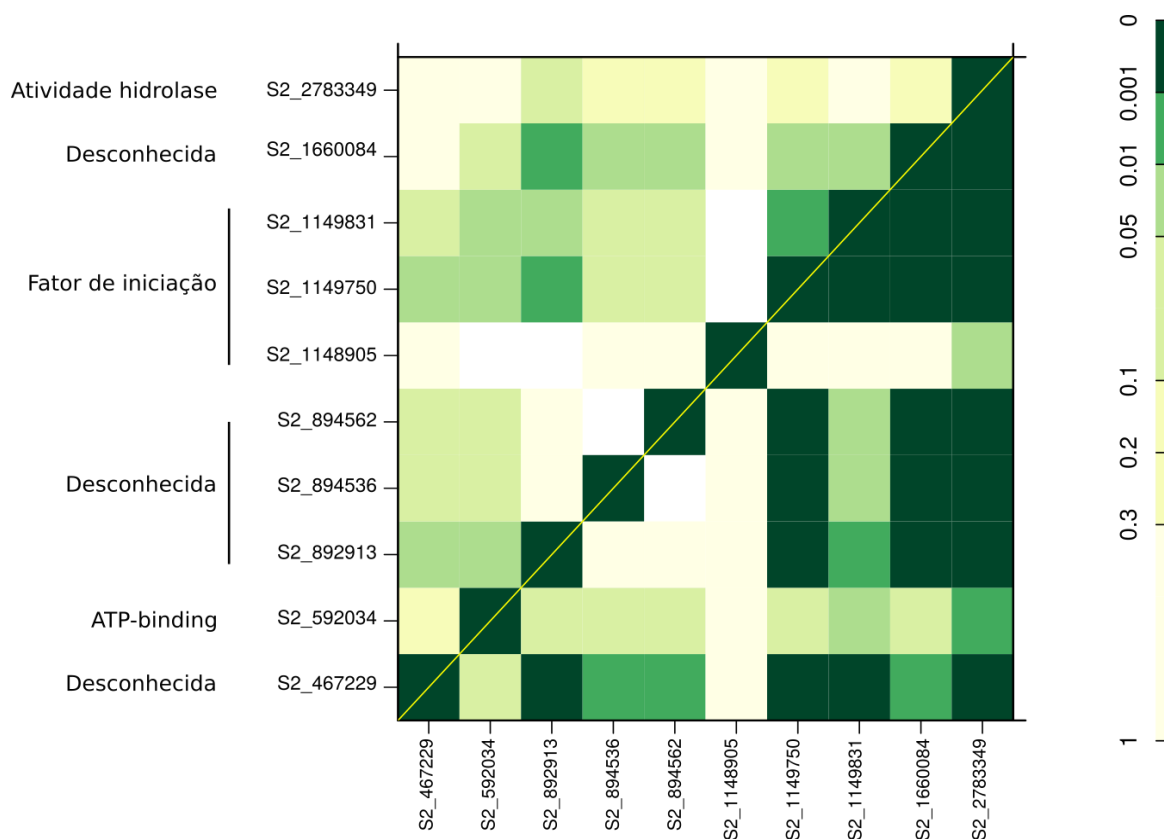


Figura 6: Interação SNP-SNP sob o modelo recessivo. O triângulo superior contém os p-valores para o teste de razão log-verossimilhança (LRT) de interação (epistasia). O triângulo inferior contém os p-valores para o teste LRT comparando a probabilidade aditiva de dois SNPs em relação ao melhor modelo *single-SNP*. A diagonal contém os p-valores do LRT para o efeito bruto de cada SNP. As diferentes tonalidades indicam os níveis de significância estatística. A atividade dos genes co-localizados aos SNPs está descrita a esquerda.

DISCUSSÃO

Associação genômica ampla

A análise GWAS relaciona polimorfismos genéticos a um fenótipo de interesse. Diferentemente da análise de mapeamento biparental dependente da variação entre dois parentais, as análises de GWAS analisam uma ampla faixa de variação natural aproveitando dos eventos de recombinação da população e do desequilíbrio de ligação para identificar mutações associadas a características fenotípicas (BURGHARDT; YOUNG; TIFFIN, 2017).

Desta forma, o desequilíbrio de ligação é um dos primeiros fatores que devem ser levados em consideração numa análise GWAS determinado a densidade de cobertura necessária para identificação de marcador-fenótipo (FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003). Uma densidade mais elevada de marcadores é necessária para obter marcadores em desequilíbrio de ligação de *loci* funcionais quando o LD decai rapidamente (VOS et al., 2017). Nosso estudo demonstrou um rápido decaimento do LD genômico de 100pb com $r^2=0,20$ na população composta por genótipos *waxy* e *não-waxy*. Outros estudos com dados GBS em mandioca indicaram o decaimento de 1,32 kb com $r^2 = 0,20$ e 3,21 kb em $r^2 = 0,10$ ao analisarem linhagens parciais (ESUMA et al., 2016), e 15 kb com $r^2 = 0,20$ no germoplasma de mandioca do Brasil (BRITO; OLIVEIRA; OLIVEIRA, 2017). Outro estudo recente demonstrou decaimento do LD com $r^2 = 0,1$ em 3kb, ao utilizarem dados de sequenciamento completo de indivíduos de origens diversas, como variedades locais, híbridos e variedades melhoradas da América Latina, Leste e Oeste da África e Ásia (RAMU et al., 2017). Cabe salientar que o padrão de decaimento do LD são específicos da população em estudo (RAFALSKI; MORGANTE, 2004). No nosso trabalho temos indivíduos do banco de germoplasma com pedigree desconhecido e estrutura populacional heterogênea (Figura 3) e indivíduos endogâmicos que explicam o rápido decaimento do LD.

Outro ponto crítico da GWAS é a estratificação da população e o grau de parentesco das amostras que pode gerar um viés na análise devido ao *background* genético (VILHJÁLMSSON; NORDBORG, 2012). Com o intuito de que SNPs relacionados a estrutura da população e o parentesco não fossem relacionados ao fenótipo a matriz Q (estrutura da população) e K (parentesco genômico) foram incorporadas ao modelo. A subdivisão em quatro grupos, que melhor explicaram a estrutura populacional pelo fastStructure foi adicionado como covariável na análise

GWAS e a matriz de parentesco foi adicionada como componente de variância poligênica aditiva (JIANG; MBATCHOU; MCPPEEK, 2015; JIANG; ZHONG; MCPPEEK, 2016) de forma a reduzir a ocorrência de falsos positivos. Entretanto, a análise QQ-plot evidenciou uma segregação precoce entre os p-valores esperados e observados, indicando que muitos deles são mais significativos do que o esperado sob a hipótese nula. Este resultado raramente é devido a verdadeiros positivos, mais sim devido à estratificação da população e diferenças sistemáticas nas frequências alélicas entre as subpopulações investigadas, de modo que um grande número de p-valores são menores que o esperado a partir do acaso (EHRET, 2010). Porém, isto não significa que os SNPs determinados não sejam úteis na discriminação do fenótipo em etapas subsequentes de validação dos marcadores identificados são necessárias.

Outro ponto chave da análise GWAS para amido *waxy* foi utilizar modelos adequados para a condição binária do fenótipo. Um fenótipo binário assume dois valores possíveis, $Y_i \in \{0,1\}$ e, portanto, tem sua média restrita entre 0 e 1 e admite uma relação particular entre sua média e variância especificada por $\sigma^2_i = \mu(1 - \mu_i)$ (JIANG; MBATCHOU; MCPPEEK, 2015). O CARAT é baseado em uma estrutura de quase-probabilidade de efeitos mistos, que modela efeitos covariáveis e explica a dependência da variância da média. Além disso, o modelo CARAT realiza uma análise retrospectiva na qual o fenótipo é tratado como variável independente e a matriz genotípica e as covariáveis como variável dependente. Desta forma, o modelo tenta explicar a partir do fenótipo as frequências esperadas do genótipo e das covariáveis sendo menos sensível à especificação incorreta do fenótipo. Numa abordagem prospectiva a matriz genotípica e as covariáveis são as variáveis dependentes e o fenótipo a variável independente, ou seja, dada uma combinação de genótipos e covariáveis o modelo tenta especificar qual fenótipo é mais provável. (JIANG; ZHONG; MCPPEEK, 2016). Mesmo sendo um método recente, o CARAT tem sido utilizado com sucesso em alguns estudos. Um exemplo é o trabalho de MCCLURE et al. (2018), que utilizou o modelo CARAT para identificação de regiões genômicas associadas à resistência a sarna da macieira, com eficiência comparável ao modelo Modelo Linear Misto - MLM. O CARAT também apresentou excelente performance em dados referentes à ausência e presença da doença de Crohn em humanos quando comparada a outros métodos comumente utilizados para avaliações binárias (JIANG; ZHONG; MCPPEEK, 2016).

SNPs associados a genes da rota metabólica do amido

A arquitetura genética do fenótipo *waxy* em mandioca está associada a poucos *locus* no cromossomo dois. Dez SNPs distribuídos a uma distância de 2316kpb, envolvendo os cinco genes conhecidos e cinco de função desconhecida foram capazes de discriminar os fenótipos *waxy* e não-*waxy* na mandioca. De modo geral, SNPs significativos encontrados em estudos de associação genômica geralmente não são causais, de tal modo que os polimorfismos associados estão em desequilíbrio com a variante causal, ao invés de terem uma função biológica relacionada (KORTE; FARLOW, 2013). Como os SNPs significativos encontrados neste trabalho estão localizados em genes e com função diretamente relacionada a expressão gênica e a degradação de amido, análises diretas sobre como estes genes exercem efeito sob o fenótipo são encorajadas.

Três SNPs identificados estão relacionados a um mesmo fator de iniciação de tradução gênica (IF-2), que é um complexo proteico com três subunidades não idênticas (α , β e γ) sendo utilizado para a tradução de praticamente todos os transcritos de mRNA (POULIN; SONENBERG, 2013). Já as proteínas ABC correspondem a uma superfamília de proteínas onde a maioria realizam o transporte de compostos via membrana com exceção das subfamílias “E” e “F”. Em plantas, a subfamília ABCF relacionada ao SNP significativo S2_592034 não foi caracterizada funcionalmente, mas em leveduras e humanos contribuem para o controle da tradução por meio da ativação da quinase eIF-2a (VERRIER et al., 2008). Em mandioca até o presente momento não foram encontrados trabalhos específicos sobre os fatores de iniciação e as proteínas ABC relacionadas a cultura. Como estes genes estão diretamente correlacionados a transcrição/tradução, mutações nestes podem exercer desordens sob os genes que exercem efeito e novos estudos devem ser realizados para confirmação destas funções.

O SNP (S2_2783349) está co-localizado no gene Beta-frutofuranosidase (Manes.02G035900) que realiza a hidrólise da sacarose em glicose e frutose nas plantas. A beta-frutofuranosidase atua no processo de degradação de sacarose III (invertase da sacarose) (Mesculenta PWY-621) e degradação de sacarose V (α -glucosidase da sacarose) (Mesculenta PWY66-373) em mandioca (dados Phytomine). A atividade da invertase ácida vacuolar está relacionada à hidrólise da sacarose e seu armazenamento pode desempenhar um papel importante no acúmulo de amido (YAO et al., 2014, 2015).

Também foram identificados quais genes relacionados a rota metabólica do amido estão próximos aos SNPs significativos pela GWAS, uma vez que é possível que os genes mais próximos dos polimorfismos identificados sejam os genes com efeito principal no fenótipo. Dois genes foram encontrados próximos as variantes detectadas pelas GWAS: GBSSI (117671-121941). e Beta-amilase (681156 a 684515). O gene GBSSI é uma sintase ligada ao alongamento da amilose e a ausência desta enzima é responsável pelo fenótipo *waxy* em diferentes culturas amiláceas como arroz (KHARABIAN-MASOULEH et al., 2012; LARKIN; PARK, 2003), milho (XIAOYANG et al., 2017) e trigo (MONARI et al., 2005; SAITO; NAKAMURA, 2005; ZHANG et al., 2017). Em mandioca mutantes *waxy* foram induzidos por modificações na expressão deste gene via transgenia (KOEHORST-VAN PUTTEN et al., 2012; RAEMAKERS et al., 2005; ZHAO et al., 2011) e uma fonte natural, o clone AM 206-5, foi identificada pelo Centro Internacional de Agricultura Tropical (CIAT) por meio de uma série de autofecundações (CEBALLOS et al., 2007). Posteriormente o sequenciamento do gene *GBSSI* proveniente de quatro progênies com amido *waxy* e não-*waxy* derivados do cruzamento entre a fonte original AM206-5 e variedades comerciais identificou potenciais mutações funcionais para seleção precoce desta característica (AIEMNAKA et al., 2012).

Diferentemente da função da GBSSI, as enzimas beta-amilase (1,4- α -D-glucanglucanohidrolase) e alfa-amilase (1,4- α -D-glucanglucanohidrolase) tem como função a degradação do amido. Estudos anteriores demonstraram alta expressão gênica da alfa-amilase nas raízes de reserva, e da beta-amilase em folhas e raízes de armazenamento (LENGBAMROUNG et al., 2005). Estes mesmos autores também identificaram múltiplas cópias da beta-amilase na mandioca, cuja localização no mapa genético atual da mandioca refere-se aos cromossomos 3, 6, 8, 12, 13, 14 e 15 (BREDESON et al., 2016).

O presente trabalho também abordou interações entre os SNPs significativos encontrados pela GWAS, sendo detectada um potencial efeito epistático e aditivo entre os SNPs em relação ao fenótipo que precisa ser melhor investigado em trabalhos futuros envolvendo fontes alternativas do gene *waxy*, assim como populações segregantes de maior tamanho. Provavelmente o efeito individual de cada mutação nos genes relacionados podem de forma conjunta resultar em fenótipos *waxy*.

Perspectivas futuras para exploração do fenótipo waxy

Os resultados encontrados dão um passo importante para a compreensão da arquitetura genética do fenótipo *waxy* em mandioca e abrem caminho para investigações subseqüentes visando estudo de variantes alélicas e de genes envolvidos no controle desta característica. Embora uma deleção de uma base no éxon seis do gene *GBSSI* tenha sido identificada como causal para o fenótipo *waxy* (AIEMNAKA et al., 2012), nosso trabalho evidencia outras regiões polimórficas relacionadas ao amido *waxy* que devem ser melhor investigadas.

Os polimorfismos detectados no nosso estudo podem direcionar novas pesquisas e revelar processos biológicos direta e indiretamente associados ao fenótipo *waxy*. Para isso, será necessário identificar a função gênica dos cinco SNPs co-localizados em regiões proteicas de função desconhecida e a relação entre os genes já descritos ao fenótipo. Estudos adicionais para determinar os mecanismos e elementos reguladores da expressão da *GBSSI* poderão elucidar as interações entre o *GBSSI* e outras proteínas conhecidas ou desconhecidas envolvidas na biossíntese do amido e mais especificamente da amilose (SEUNG et al., 2015, 2017). Existem poucas informações sobre o potencial de regulação pós-tradução nas enzimas do metabolismo do amido e evidências sugerem que o *GBSSI* é regulado no processo de pós-tradução (GRIMAUD et al., 2008; XING et al., 2016).

Além disso, existem importantes avanços na implementação prática da seleção genômica em mandioca para aumento da resistência a doenças, da produtividade e qualidade de raízes (WOLFE et al., 2016). Portanto, em um futuro próximo a seleção genômica poderá ser aplicada ao desenvolvimento de novas variedades *waxy* de mandioca para aumento da produtividade de raízes e de amido, tendo em vista que estudos recentes demonstram menor teor de matéria seca nas raízes nos genótipos *waxy* (KARLSTRÖM et al., 2016). Contudo, como se trata de um fenótipo recessivo, é necessário produzir as sementes botânicas dos cruzamentos planejados, realizar as avaliações de campo e a multiplicação do material selecionado que levam de 9 a 11 anos para obtenção de plantas com o fenótipo *waxy* recessivo. Por isso, a localização das regiões genômicas em LD com o fenótipo *waxy* pode facilitar a seleção conjunta dos indivíduos nas populações segregantes que possuam o fenótipo *waxy* de forma precoce (na fase de *seedlings*) e ainda permita estimar os *Genomic Estimated Breeding Values (GEBV)* dos mais promissores para produtividade de raízes e amido, considerando uma mesma genotipagem.

REFERÊNCIAS

- AIEMNAKA, P.; WONGKAEW, A.; CHANTHAWORN, J.; NAGASHIMA, S. K.; BOONMA, S.; AUTHAPUN, J.; JENWEERAWAT, S.; KONGSILA, P.; KITTIPADAKUL, P.; NAKASATHIEN, S.; SREEWONGCHAI, T.; WANNARAT, W.; VICHUKIT, V.; LÓPEZ-LAVALLE, L. A. B.; CEBALLOS, H.; ROJANARIDPICHED, C.; PHUMICHA, C. Molecular characterization of a spontaneous waxy starch mutation in cassava. **Crop Science**, v. 52, n. 5, p. 2121–2130, 2012.
- 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**, v. 23, n. 19, p. 2633–2635, 2007.
- BREDESON, J. V.; LYONS, J. B.; PROCHNIK, S. E.; WU, G. A.; HA, C. M.; EDSINGER-GONZALES, E.; GRIMWOOD, J.; SCHMUTZ, J.; RABBI, I. Y.; EGESI, C.; NAULUVULA, P.; LEBOT, V.; NDUNGURU, J.; MKAMILO, G.; BART, R. S.; SETTER, T. L.; GLEADOW, R. M.; KULAKOW, P.; FERGUSON, M. E.; ROUNSLEY, S.; ROKHSAR, D. S. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. **Nature biotechnology**, v. 34, n. 5, p. 562–570, 2016.
- BRITO, A. C.; OLIVEIRA, S. A. S.; OLIVEIRA, E. J. Genome-wide association study for resistance to cassava root rot. **The Journal of Agricultural Science**, v. 155, n. 09, p. 1424–1441, 2017.
- BROWNING, B. L.; BROWNING, S. R. Improving the accuracy and efficiency of identity-by-descent detection in population data. **Genetics**, v. 194, n. 2, p. 459–471, 2013.
- BROWNING, B. L.; BROWNING, S. R. Genotype Imputation with Millions of Reference Samples. **American journal of human genetics**, v. 98, n. 1, p. 116–26, 2016.
- BURGHARDT, L. T.; YOUNG, N. D.; TIFFIN, P. A Guide to Genome-Wide Association Mapping in Plants. In: **Current Protocols in Plant Biology**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2017. p. 22–38.
- CEBALLOS, H.; IGLESIAS, C. a.; PÉREZ, J. C.; DIXON, A. G. O. Cassava

breeding: Opportunities and challenges. **Plant Molecular Biology**, v. 56, n. 4, p. 503–516, 2004.

CEBALLOS, H.; KULAKOW, P.; HERSHEY, C. Cassava Breeding: Current Status, Bottlenecks and the Potential of Biotechnology Tools. **Tropical Plant Biology**, v. 5, n. 1, p. 73–87, 2012. Disponível em: <<http://link.springer.com/10.1007/s12042-012-9094-9>>

CEBALLOS, H.; SÁNCHEZ, T.; MORANTE, N.; FREGENE, M.; DUFOUR, D.; SMITH, A. M.; DENYER, K.; PÉREZ, J. C.; CALLE, F.; MESTRES, C. Discovery of an amylose-free starch mutant in cassava (*Manihot esculenta* Crantz). **Journal of Agricultural and Food Chemistry**, v. 55, n. 18, p. 7469–7476, 2007.

DENYER, K.; JOHNSON, P.; ZEEMAN, S.; SMITH, A. M. The control of amylose synthesis. **Journal of Plant Physiology**, v. 158, n. 4, p. 479–487, 2001.

DESCHAMPS, P.; HAFERKAMP, I.; D'HULST, C.; NEUHAUS, H. E.; BALL, S. G. The relocation of starch metabolism to chloroplasts: when, why and how. **Trends in Plant Science**, v. 13, n. 11, p. 574–582, 2008.

DOBO, M.; AYRES, N.; WALKER, G.; PARK, W. D. Polymorphism in the GBSS gene affects amylose content in US and European rice germplasm. **Journal of Cereal Science**, v. 52, n. 3, p. 450–456, 2010.

DOYLE, J.; DOYLE, J. Isolation of plant DNA from fresh tissue. **Focus**, v. 12, p. 13–15, 1990.

EHRET, G. B. Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension. **Current hypertension reports**, v. 12, n. 1, p. 17–25, 2010.

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**, v. 6, n. 5, p. 1–10, 2011.

ESUMA, W.; HERSELMAN, L.; LABUSCHAGNE, M. T.; RAMU, P.; 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.

FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER, E. S. Structure of Linkage disequilibrium in plants. **Annual Review of Plant Biology**, v. 54, n. 1, p. 357–374, 2003.

FUJITA, N.; YOSHIDA, M.; KONDO, T.; SAITO, K.; UTSUMI, Y.; TOKUNAGA, T.; NISHI, A.; SATOH, H.; PARK, J.-H.; JANE, J.-L.; MIYAO, A.; HIROCHIKA, H.; NAKAMURA, Y. Characterization of SSIIIa-deficient mutants of rice: The function of SSIIIa and pleiotropic effects by SSIIIa deficiency in the Rice endosperm. **Plant Physiology**, v. 144, n. 4, p. 2009–2023, 2007.

GEERA, B. P.; NELSON, J. E.; SOUZA, E.; HUBER, K. C. Granule bound starch synthase I (GBSSI) gene effects related to soft wheat flour/starch characteristics and properties. **Cereal Chemistry**, v. 83, n. 5, p. 544–550, 2006.

GONZÁLEZ, J. R.; ARMENGOL, L.; GUINÓ, E.; SOLÉ, X.; MORENO, V. **SNPassoc: SNPs-based whole genome association studies**, 2014. Disponível em: <<https://cran.r-project.org/package=SNPassoc>>

GONZÁLEZ, J. R.; ARMENGOL, L.; SOLÉ, X.; GUINÓ, E.; MERCADER, J. M.; ESTIVILL, X.; MORENO, V. SNPassoc: An R package to perform whole genome association studies. **Bioinformatics**, v. 23, n. 5, p. 644–645, 2007.

GOODSTEIN, D. M.; SHU, S.; HOWSON, R.; NEUPANE, R.; HAYES, R. D.; FAZO, J.; MITROS, T.; DIRKS, W.; HELLSTEN, U.; PUTNAM, N.; ROKHSAR, D. S. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, n. D1, p. D1178–D1186, 2012.

GRIMAUD, F.; ROGNIAUX, H.; JAMES, M. G.; MYERS, A. M.; PLANCHOT, V. Proteome and phosphoproteome analysis of starch granule-associated proteins from normal maize and mutants affected in starch biosynthesis. **Journal of Experimental Botany**, v. 59, n. 12, p. 3395–3406, 2008.

JIANG, D.; MBATCHOU, J.; MCPEEK, M. S. Retrospective association analysis of binary traits: overcoming some limitations of the additive polygenic model. **Human Heredity**, v. 80, n. 4, p. 187–195, 2015.

JIANG, D.; ZHONG, S.; MCPEEK, M. S. Retrospective binary-trait association test elucidates genetic architecture of Crohn disease. **American Journal of Human Genetics**, v. 98, n. 2, p. 243–255, 2016.

KARLSTRÖM, A.; CALLE, F.; SALAZAR, S.; MORANTE, N.; DUFOUR, D.; CEBALLOS, H. Biological implications in Cassava for the production of amylose-free starch: Impact on root yield and related traits. **Frontiers in Plant Science**, v. 7, p. 1–8, 2016.

KAYONDO, S. I.; PINO DEL CARPIO, D.; LOZANO, R.; OZIMATI, A.; WOLFE, M.; BAGUMA, Y.; GRACEN, V.; OFFEI, S.; FERGUSON, M.; KAWUKI, R.; JANNINK, J.-L. Genome-wide association mapping and genomic prediction for CBSD resistance in *Manihot esculenta*. **Scientific Reports**, v. 8, n. 1, p. 1549, 2018.

KHARABIAN-MASOULEH, A.; WATERS, D. L. E.; REINKE, R. F.; WARD, R.; HENRY, R. J. SNP in starch biosynthesis genes associated with nutritional and functional properties of rice. **Scientific reports**, v. 2, p. 557, 2012.

KOEHORST-VAN PUTTEN, H. J. J.; SUDARMONOWATI, E.; HERMAN, M.; PEREIRA-BERTRAM, I. J.; WOLTERS, A. M. A.; MEIMA, H.; DE VETTEN, N.; RAEMAKERS, C. J. J. M.; VISSER, R. G. F. Field testing and exploitation of genetically modified cassava with low-amylose or amylose-free starch in Indonesia. **Transgenic Research**, v. 21, n. 1, p. 39–50, 2012.

KORTE, A.; FARLOW, A. The advantages and limitations of trait analysis with GWAS: a review. **Plant Methods**, v. 9, n. 1, p. 29, 2013.

LARKIN, P. D.; PARK, W. D. Association of waxy gene single nucleotide polymorphisms with starch characteristics in rice (*Oryza sativa* L.). **Molecular Breeding**, v. 12, n. 4, p. 335–339, 2003.

LENGBAMROUNG, P.; VICHUKIT, V.; VISSER, R. G. F.; NAKASATHIEN, S. Cloning and Molecular Characterization of a - and b -Amylase Genes from Cassava (*Manihot esculenta* Crantz). **Natural science**, v. 39, p. 446–454, 2005.

LIPKA, A. E.; TIAN, F.; WANG, Q.; PEIFFER, J.; LI, M.; BRADBURY, P. J.; GORE, M. A.; BUCKLER, E. S.; ZHANG, Z. GAPIT: genome association and prediction integrated tool. **Bioinformatics**, v. 28, n. 18, p. 2397–2399, 2012. Disponível em:

<<http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/bts444>>

MONARI, A. M.; SIMEONE, M. C.; URBANO, M.; MARGIOTTA, B.; LAFIANDRA, D. Molecular characterization of new waxy mutants identified in bread and durum wheat. **Theoretical and Applied Genetics**, v. 110, n. 8, p. 1481–1489, 2005.

MOORTHY, S. N. Physicochemical and Functional Properties of Tropical Tuber Starches: A Review. **Starch - Stärke**, v. 54, n. 12, p. 559–592, 2002.

PÉREZ, S.; BERTOFT, E. The molecular structures of starch components and their contribution to the architecture of starch granules: A comprehensive review. **Starch/Staerke**, v. 62, p. 389–420, 2010.

PFISTER, B.; ZEEMAN, S. C. Formation of starch in plant cells. **Cellular and Molecular Life Sciences**, v. 73, n. 14, p. 2781–2807, 2016.

POULIN, F.; SONENBERG, N. **Mechanism of Translation Initiation in Eukaryotes**. Austin, EUA: Landes Bioscience, 2013. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK6597/>>

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I. W.; DALY, M. J.; SHAM, P. C. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**, 2018. Disponível em: <<https://www.r-project.org/>>

RABBI, I. Y.; UDOH, L. I.; WOLFE, M.; PARKES, E. Y.; GEDIL, M. A.; DIXON, A.; RAMU, P.; JANNINK, J.-L.; KULAKOW, P. Genome-Wide Association Mapping of Correlated Traits in Cassava: Dry Matter and Total Carotenoid Content. **The Plant Genome**, v. 10, n. 3, p. 1–14, 2017.

RAEMAKERS, K.; SCHREUDER, M.; SUURS, L.; FURRER-VERHORST, H.; VINCKEN, J.-P.; DE VETTEN, N.; JACOBSEN, E.; VISSER, R. G. F. Improved Cassava Starch by Antisense Inhibition of Granule-bound Starch Synthase I. **Molecular Breeding**, v. 16, n. 2, p. 163–172, 2005.

RAFALSKI, A.; MORGANTE, M. Corn and humans: Recombination and linkage disequilibrium in two genomes of similar size. **Trends in Genetics**, v. 20, n. 2, p. 103–111, 2004.

RAHEEM, D.; CHUKWUMA, C. Foods from cassava and their relevance to Nigeria and other African countries. **Agriculture and Human Values**, p. 383–390, 2001.

RAJ, A.; STEPHENS, M.; PRITCHARD, J. K. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. **Genetics**, v. 197, n. 2, p. 573–589, 2014.

RAMU, P.; ESUMA, W.; KAWUKI, R.; RABBI, I. Y.; EGESI, C.; BREDESON, J. V.; BART, R. S.; VERMA, J.; BUCKLER, E. S.; LU, F. Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. **Nature Genetics**, v. 49, n. 6, p. 959–963, 2017.

SAITO, M.; NAKAMURA, T. Two point mutations identified in emmer wheat generate null Wx-A1 alleles. **Theoretical and Applied Genetics**, v. 110, n. 2, p. 276–282, 2005.

SÁNCHEZ, T.; DUFOUR, D.; MORENO, I. X.; CEBALLOS, H. Comparison of pasting and gel stabilities of waxy and normal starches from potato, maize, and rice with those of a novel waxy cassava starch under thermal, chemical, and mechanical stress. **Journal of Agricultural and Food Chemistry**, v. 58, n. 8, p. 5093–5099, 2010.

SEUNG, D.; BOUDET, J.; MONROE, J.; SCHREIER, T. B.; DAVID, L. C.; ABT, M.; LU, K.; ZANELLA, M.; ZEEMAN, S. C. Homologs of Protein Targeting to Starch control starch granule initiation in Arabidopsis leaves. **The Plant Cell**, v. 29, n. 7, p. 1657 LP-1677, 2017.

SEUNG, D.; SOYK, S.; COIRO, M.; MAIER, B. A.; EICKE, S.; ZEEMAN, S. C. Protein Targeting to Starch is required for localising Granule-bound starch synthase to Starch Granules and for normal amylose synthesis in Arabidopsis. **PLOS Biology**, v. 13, n. 2, p. e1002080, 2015.

SHIN, J.-H.; BLAY, S.; MCNENEY, B.; GRAHAM, J. LDheatmap: An R function for graphical display of pairwise Linkage Disequilibria between Single Nucleotide

Polymorphisms. **Journal of Statistical Software**, v. 16, p. 1–10, 2006.

TANG, Y.; LIU, X.; WANG, J.; LI, M.; WANG, Q.; TIAN, F.; SU, Z.; PAN, Y.; LIU, D.; LIPKA, A. E.; BUCKLER, E. S.; ZHANG, Z. GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. **The Plant Genome**, v. 9, n. 2, 2016.

TETLOW, I. J.; MORELL, M. K.; EMES, M. J. Recent developments in understanding the regulation of starch metabolism in higher plants. **Journal of Experimental Botany**, v. 55, n. 406, p. 2131–2145, 2004.

TURNER, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. **Journal of Open Source Software**, v. 3, n. 25, p. 731, 2018.

VANRADEN, P. M. Efficient Methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414–4423, 2008.

VERRIER, P. J.; BIRD, D.; BURLA, B.; DASSA, E.; FORESTIER, C.; GEISLER, M.; KLEIN, M.; KOLUKISA OGLU, Ü.; LEE, Y.; MARTINOIA, E.; MURPHY, A.; REA, P. A.; SAMUELS, L.; SCHULZ, B.; SPALDING, E. J.; YAZAKI, K.; THEODOULOU, F. L. Plant ABC proteins - a unified nomenclature and updated inventory. **Trends in Plant Science**, v. 13, n. 4, p. 151–159, 2008.

VILHJÁLMSSON, B. J.; NORDBORG, M. The nature of confounding in genome-wide association studies. **Nature Reviews Genetics**, v. 14, n. 1, p. 1–2, 2012.

VOS, P. G.; PAULO, M. J.; VOORRIPS, R. E.; VISSER, R. G. F.; VAN ECK, H. J.; VAN EEUWIJK, F. A. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. **Theoretical and Applied Genetics**, v. 130, n. 1, p. 123–135, 2017.

WEISBERG, S. **Applied linear regression**. [s.l.] : John Wiley & Sons, 2005. v. 528

WOLFE, M. D.; RABBI, I. Y.; EGESI, C.; HAMBLIN, M.; KAWUKI, R.; KULAKOW, P.; LOZANO, R.; CARPIO, D. P. Del; RAMU, P.; JANNINK, J.-L. Genome-Wide Association and Prediction Reveals Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement. **The Plant Genome**, v. 9, n. 2, 2016.

XIAOYANG, W.; DAN, C.; YUQING, L.; WEIHUA, L.; XINMING, Y.; XIUQUAN, L.; JUAN, D.; LIHUI, L. Molecular characteristics of two new waxy mutations in China waxy maize. **Molecular Breeding**, v. 37, n. 3, 2017.

XING, S.; MENG, X.; ZHOU, L.; MUJAHID, H.; ZHAO, C.; ZHANG, Y.; WANG, C.; PENG, Z. Proteome profile of starch granules purified from rice (*Oryza sativa*) endosperm. **PLoS ONE**, v. 11, n. 12, p. 1–20, 2016.

YAMAMORI, M.; QUYNH, N. T. Differential effects of Wx-A1, -B1 and -D1 protein deficiencies on apparent amylose content and starch pasting properties in common wheat. **TAG Theoretical and Applied Genetics**, v. 100, n. 1, p. 32–38, 2000.

YAO, Y.; GENG, M.-T.; WU, X.-H.; LIU, J.; LI, R.-M.; HU, X.-W.; GUO, J.-C. Genome-Wide Identification, Expression, and Activity Analysis of Alkaline/Neutral Invertase Gene Family from Cassava (*Manihot esculenta* Crantz). **Plant Molecular Biology Reporter**, v. 33, n. 2, p. 304–315, 2015.

YAO, Y.; WU, X. H.; GENG, M. T.; LI, R. M.; LIU, J.; HU, X. W.; GUO, J. C. Cloning, 3D modeling and expression analysis of three vacuolar invertase genes from cassava (*Manihot Esculenta* Crantz). **Molecules**, v. 19, n. 5, p. 6228–6245, 2014.

ZHANG, L.; CHEN, H.; LUO, M.; ZHANG, X.; DENG, M.; MA, J.; QI, P.; WANG, J.; CHEN, G.; LIU, Y.; PU, Z.; LI, W.; LAN, X.; WEI, Y.; ZHENG, Y.; JIANG, Q. Transposon insertion resulted in the silencing of Wx-B1n in Chinese wheat landraces. **Theoretical and Applied Genetics**, v. 130, n. 6, p. 1331–1331, 2017.

ZHAO, S.-S.; DUFOUR, D.; SÁNCHEZ, T.; CEBALLOS, H.; ZHANG, P. Development of waxy cassava with different Biological and physico-chemical characteristics of starches for industrial applications. **Biotechnology and Bioengineering**, v. 108, n. 8, p. 1925–1935, 2011.

MATERIAL SUPLEMENTAR

Tabela S1: Relação de genótipos utilizados nos estudos de associação genômica ampla (GWAS) do Banco de Germoplasma de Mandioca.

Amido waxy			
6383_01wx	7694_03wx	7736_03wx	7738_01wx
6413_01wx	7697_03wx	7736_05wx	7745_04wx
6421_05wx	7698_14wx	7736_07wx	7745_05wx
6428_01wx	7698_16wx	7737_01wx	7745_06wx
6430_01wx	7734_01wx	7737_02wx	8109_03wx
6437_01wx	7734_02wx	7737_06wx	8109_04wx
6477_02wx	7734_06wx	7737_07wx	8109_05wx
7690_08wx	7734_07wx	7737_08wx	
Amido não waxy			
Água_Morna	BGM0186	BGM0699	BGM0930
Aimpim_Brasil	BGM0205	BGM0776	BGM0931
Aipim_497	BGM0207	BGM0780	BGM0932
Aipim_Abacate	BGM0209	BGM0791	BGM0934
Aipim_Brasilia	BGM0213	BGM0800	BGM0935
Aipim_Manteiga	BGM0215	BGM0805	BGM0936
Aipim_Ze_Marcos	BGM0216	BGM0806	BGM0942
Alagoana_363	BGM0217	BGM0808	BGM0943
Alagoana_381	BGM0236	BGM0817	BGM0945
Amansa_Burro	BGM0246	BGM0826	BGM0946
Amarelol	BGM0252	BGM0829	BGM0947
AmarelolI	BGM0255	BGM0830	BGM0948
Amparodospobres	BGM0265	BGM0834	BGM0949
Baianinha	BGM0267	BGM0847	BGM0951
BGM0017	BGM0270	BGM0854	BGM0951
BGM0022TB	BGM0271	BGM0855	BGM0952
BGM0022TR	BGM0290	BGM0863	BGM0954
BGM0024	BGM0315	BGM0865	BGM0957
BGM0038	BGM0320	BGM0869	BGM0961
BGM0072	BGM0374	BGM0870	BGM0966
BGM0111	BGM0424	BGM0872	BGM0971
BGM0115TB	BGM0450	BGM0891	BGM0972
BGM0115TR	BGM0476	BGM0901	BGM0975
BGM0125	BGM0505CMe	BGM0907	BGM0976
BGM0132	BGM0505CP	BGM0914	BGM0979
BGM0137	BGM0584	BGM0918	BGM0981
BGM0151	BGM0609	BGM0919	BGM0983
BGM0152	BGM0649	BGM0922	BGM0984
BGM0155	BGM0650	BGM0928	BGM0994
BGM0168	BGM0652	BGM0929	BGM1017

Tabela S1: continuação...

Amido não waxy			
BGM1034	BGM1477	BGM1879	BGM2102
BGM1064	BGM1481	BGM1992	BGM2103
BGM1097	BGM1484	BGM2003	BGM2104
BGM1124	BGM1486	BGM2004	BGM2105
BGM1147	BGM1492	BGM2019	BGM2110
BGM1161	BGM1497	BGM2035	BGM2121
BGM1176	BGM1502	BGM2041	BGM2123
BGM1212	BGM1503	BGM2042	BGM2140
BGM1226	BGM1508	BGM2043_1	BGM2146
BGM1253	BGM1515	BGM2043_2	BGM2149
BGM1304	BGM1517	BGM2044	BGM2234
BGM1311	BGM1528	BGM2047	BGM2235
BGM1316	BGM1539	BGM2048	Boazinha_359
BGM1326	BGM1544	BGM2051	Branca_208
BGM1343	BGM1546	BGM2052	Branca_Curta_5
BGM1361	BGM1547	BGM2053	Branca_SC
BGM1393	BGM1550	BGM2054	Branco_Aipim_199
BGM1398	BGM1571	BGM2055	Branco_Aipim_375
BGM1400	BGM1573	BGM2057	Branquinha
BGM1402	BGM1580	BGM2061	Branquinha_405
BGM1403	BGM1627	BGM2063	Brasileira_Uruas_106
BGM1407	BGM1632	BGM2064	BRS 396
BGM1409	BGM1668	BGM2065	BRS Caipira
BGM1413	BGM1680	BGM2066	BRS Kiriris
BGM1415	BGM1683	BGM2067	BRS Tapioqueira
BGM1420	BGM1685	BGM2070	Cacau
BGM1422	BGM1691	BGM2071	Cachimbo
BGM1424	BGM1695	BGM2072	Canário
BGM1425	BGM1717	BGM2073	Canela_de_Jacu_250
BGM1426	BGM1721	BGM2074	Cigana428
BGM1427	BGM1722	BGM2076	CiganaPreta354
BGM1436	BGM1728	BGM2080	Clone_RI
BGM1438	BGM1748	BGM2082	Correntão
BGM1443	BGM1755	BGM2083	Craibeirinha_345
BGM1448	BGM1759	BGM2084	Cramuquem_377
BGM1453	BGM1762	BGM2086	Crateus_433
BGM1458	BGM1782	BGM2091	Cria_Menino_372
BGM1465	BGM1784	BGM2095	Dimmy01
BGM1476	BGM1817	BGM2096	Dimmy02

Tabela S1: continuação...

Amido não waxy			
Dimmy03	Guela de Jacu	Mico	Platina_371
Dimmy05	Ipirá	Milagrosa	Preta_213
Dimmy08	Isabel_de_Souza_225	Mineira_339	Preta_247
Dimmy09	Jacum_394	Moura	Preta_249
Dimmy10	Joaquinzinha_201	Mutambo_Joselito	Pretinha_114
Dimmy12	Joaquinzinha_392	Nagib_ICB_307_2	Quinze_Minutos_248
Dimmy13	Joselito_A2	Nagib_ICB_388	Recife_115
Dimmy14	KU_50	Naja_338	Riqueza_215
Dimmy16	Kun	Naja_Vermelha	SC_008
Dimmy19	Lagoão	Naja_Vermelha_432	SC_009
Dimmy20	M_Caeru_766	Nass_Gig	SC_124
Dimmy21	Mane_Roque	Nega_Maluca	SC_205
Dimmy23	Manipeba_408	Nuomi	SE_01
Dimmy25	Manipeba_Prog_108	Olho_Junto	SE_02
Dimmy27	Manteiga_369	Ouro_Pão	SE_03
Dimmy28	Manzhital	Paragu_PR	SE_09
Engana_Ladrão	Manzhital_199	Pau_Branco	SI_Rosa
Enrica_Homen_212	Mata_Fome_Preta_385	Peixe	Tio_Goncalo
Folha_Fina_229	Miambão	Pioneira	TMe_14

Tabela S2: Relação de genes relacionados síntese de amido em mandioca

Síntese de amido				
Transcrito	Descrição	Chr	Início	Fim
Manes.03G133300	Enzima ramificadora de 1,4-alfa-glucana (EC 2.4.1.18)	3	22943836	22953045
Manes.05G133800	Enzima ramificadora de 1,4-alfa-glucana (EC 2.4.1.18)	5	17845753	17853349
Manes.08G022400	Enzima ramificadora de 1,4-alfa-glucana (EC 2.4.1.18)	8	2066905	2083856
Manes.08G022500	Enzima ramificadora de 1,4-alfa-glucana (EC 2.4.1.18)	8	2084505	2095963
Manes.09G059400	Enzima ramificadora de 1,4-alfa-glucana (EC 2.4.1.18)	9	7914812	7934606
Manes.14G031100	Fosfoglucomutase (cofator-glicose) (EC 5.4.2.5)	14	2550417	2560141
Manes.15G131800	Fosfoglucomutase (cofator-glicose) (EC 5.4.2.5)	15	10060554	10069139
Manes.17G081000	Fosfoglucomutase (cofator-glicose) (EC 5.4.2.5)	17	22089798	22098089
Manes.06G141300	Fosfoglucomutase (cofator-glicose) (EC 5.4.2.5)	6	24600899	24610525
Manes.10G046600	Fosfoglucomutase (dependente de alfa-D-glicose-1,6-bisfosfato) (EC 5.4.2.2)	10	4427088	4440188
Manes.17G079900	Fosfoglucomutase (dependente de alfa-D-glicose-1,6-bisfosfato) (EC 5.4.2.2)	17	22007934	22011166
Manes.02G146400	Glicose-6-fosfato isomerase (EC 5.3.1.9)	2	10872560	10880086
Manes.05G160800	Glicose-6-fosfato isomerase (EC 5.3.1.9)	5	22599759	22609213
Manes.18G060600	Glicose-6-fosfato isomerase (EC 5.3.1.9)	18	5253290	5267876
Manes.18G027800	Glicose-6-fosfato isomerase (EC 5.3.1.9)	18	2561557	2570413
Manes.01G236700	Glucose-1-fosfato adenililtransferase (EC 2.7.7.27)	1	31771894	31777649
Manes.03G182100	Glucose-1-fosfato adenililtransferase (EC 2.7.7.27)	3	26947144	26950742
Manes.05G208000	Glucose-1-fosfato adenililtransferase (EC 2.7.7.27)	5	28304177	28308021
Manes.11G085500	Glucose-1-fosfato adenililtransferase (EC 2.7.7.27)	11	11864610	11871278
Manes.12G067900	Glucose-1-fosfato adenililtransferase (EC 2.7.7.27)	12	6583306	6588233
Manes.13G058900	Glucose-1-fosfato adenililtransferase (EC 2.7.7.27)	13	6335420	6339941
Manes.15G025400	Glucose-1-fosfato adenililtransferase (EC 2.7.7.27)	15	1902603	1906248
Manes.18G019600	Glucose-1-fosfato adenililtransferase (EC 2.7.7.27)	18	1530942	1533519

Tabela S2: continuação...

Síntese de amido				
Transcrito	Descrição	Chr	Início	Fim
Manes.05G073400	Isoamilase (EC 3.2.1.68)	5	5582022	5585429
Manes.06G054000	Isoamilase (EC 3.2.1.68)	6	15559616	15574428
Manes.18G063500	Isoamilase (EC 3.2.1.68)	18	5448339	5459259
Manes.01G055700	NDP-glicose-glicosiltransferase de amido/ <i>Waxy</i> (EC 2.4.1.242)	1	15630937	15634504
Manes.02G001000	NDP-glucose -starch glucosyltransferase / <i>Waxy</i> (EC 2.4.1.242)	2	117671	121941
Manes.01G184000	Sintase do amido (EC 2.4.1.21)	1	28273525	28281524
Manes.01G091700	Sintase do amido (EC 2.4.1.21)	1	21623316	21629007
Manes.02G082800	Sintase do amido (EC 2.4.1.21)	2	6203128	6213145
Manes.02G046900	Sintase do amido (EC 2.4.1.21)	2	3606695	3614000
Manes.13G105300	Sintase do amido (EC 2.4.1.21)	13	22854164	22858114
Manes.15G118600	Sintase do amido (EC 2.4.1.21)	15	8973978	8983153
Manes.16G006900	Sintase do amido (EC 2.4.1.21)	16	682293	693880

Fonte: <https://phytozome.jgi.doe.gov/phytozome> Identificador síntese (Mesculenta PWY-622);

Descrição das enzimas por: <https://enzyme.expasy.org>

ARTIGO 3

IDENTIFICAÇÃO DE GENÓTIPOS WAXY DE MANDIOCA COM USO DO INFRAVERMELHO PRÓXIMO COM TRANSFORMADA DE FOURIER (FT-NIR) ¹

¹Artigo a ser ajustado para posterior submissão ao Comitê Editorial do periódico científico New Phytologist, em versão na língua inglesa.

Identificação de genótipos *waxy* de mandioca com uso do Infravermelho Próximo com transformada de Fourier (FT-NIR)

Resumo: A busca por ferramentas de fenotipagem em larga escala que permitam avaliação precoce e acurada das características de importância agronômica tem ganhado espaço nos diversos programas de melhoramento genético. O objetivo deste trabalho foi avaliar o potencial do infravermelho próximo com transformada de Fourier (FT-NIR) para a identificação de genótipos com amido *waxy* e validar modelos de predição para sua classificação. Foram avaliados espectros de 238 genótipos *waxy* e 208 não-*waxy* de mandioca por reflectância difusa, oriundos de cinco diferentes ambientes de cultivo. A mediana dos espectros e a análise de componentes principais foram utilizadas para observar o potencial de agrupamento dos dados, identificar genótipos discrepantes e ainda avaliar a possibilidade de componentes na classificação dos genótipos. Para classificação foram testadas cinco técnicas de reconhecimento de padrões supervisionado: *Bayesian generalized linear model* (BGLM), *high dimensional discriminant analysis* (HDDA), *partial least squares* (PLS), *parallel random forest* (PRANDF) e *support vector machines with linear kernel* (SVM). A mediana dos dados pré-processados e a análise de PCA não permitiram a discriminação dos genótipos com base na classificação do amido. Os algoritmos de SVM e BGLM apresentaram as maiores acurácias de classificação na validação cruzada (0,86 a 0,87), melhores índices de concordância, (média de 0,88 e 0,83), melhores valores de *Receiver operating characteristics* (média 0,95 e 0,94), sensibilidade (média de 0,87 e 0,85) e especificidade (0,88). Os modelos BGLM e SVM também obtiveram melhores índices na validação externa com alta precisão (0,85) e classificando corretamente 93% dos genótipos *waxy*. Desta forma, a possibilidade de realizar seleção precoce de características de raiz, por meio da seleção indireta de variáveis extraídos da folha são uma boa estratégia para o melhoramento do fenótipo *waxy*.

Palavras-chave: *Manihot esculenta* Crantz, qualidade de amido, métodos de classificação, espectroscopia.

Identification of waxy cassava genotypes using Near infrared with Fourier transform (FT-NIR)

Abstract: The search for high-throughput phenotyping tools that allow the early and accurate evaluation of the most important agronomic characteristics has gained space in several genetic improvement programs. The aim of this study was to evaluate the potential of the near infrared with Fourier transform (FT-NIR) to identify genotypes with waxy starch and to validate prediction models for their classification. We analyzed the spectra of 238 waxy and 208 non-waxy genotypes of cassava by diffuse reflectance from five different growing environments. The spectra median and principal component analysis were used to observe the potential of grouping the data, identifying discrepant genotypes and to evaluate the use of components in genotype classification. For classification, five supervised pattern recognition techniques were tested: *Bayesian generalized linear model* (BGLM), *high dimensional discriminant analysis* (HDDA), *partial least squares* (PLS), *parallel random forest* (PRANDF) and *support vector machines with linear kernel* (SVM). The median of pre-processed data and the principal component analysis (PCA) did not allow discrimination of the genotypes based on the starch classification. The SVM and BGLM algorithms showed the highest classification accuracy in cross-validation (0.86 to 0.87), better concordance rates (means 0.88 and 0.83), better values of Receiver operating characteristics (means 0.95 and 0.94), sensitivity (means 0.87 and 0.85) and specificity (0.88). The BGLM and SVM models also obtained better indexes in the external validation with high accuracy (0.85) and classifying correctly 93% of the waxy genotypes. Thus, the possibility of performing early selection of root characteristics, by the indirect selection of variables extracted from the leaf is a good strategy for the breeding of the waxy phenotype.

Keywords: *Manihot esculenta* Crantz, starch quality, classification methods, spectroscopy.

INTRODUÇÃO

O amido de mandioca (*Manihot esculenta* Crantz) possui alta viscosidade, clareza, baixo teor de lipídios de fácil gelatinização (MOORTHY, 2002) sendo o principal componente nas raízes (OLOMO; AJIBOLA, 2006). O amido de mandioca possui propriedades de pasta superiores aos amidos de outras fontes, apresentando menor sinérese quando comparado ao amido ceroso de outras plantas cultivadas (SÁNCHEZ et al., 2010), sabor neutro devido ao seu baixo teor de lipídios e proteínas, o que lhe confere vantagem em relação aos amidos de cereais para a indústria de alimentos. O amido é utilizado principalmente na indústria alimentícia, como estabilizador e em diferentes aplicações nas industriais têxtil e de papel (BREUNINGER; PIYACHOMKWAN; SRIROTH, 2009).

Novas propriedades do amido de mandioca tem sido demandas para uso na indústria, especialmente na indústria de alimentos. É neste cenário que o amido com baixo conteúdo de amilose (*waxy*) tem ganhado espaço nos programas de pesquisa de diversas instituições, sobretudo para o desenvolvimento de variedades com esta característica, bem como a exploração de suas aplicações industriais (KARLSTRÖM et al., 2016; ZHAO et al., 2011). O amido *waxy* reduz a liberação de água (sinérese) em pastas de amido durante o processo de retrogradação, evitando assim modificações na qualidade sensorial e no período de armazenamento dos alimentos (MORANTE et al., 2016; WANG et al., 2015).

Nos últimos anos, foram obtidos avanços importantes na obtenção de variedades *waxy* em mandioca via transgenia (KOEHORST-VAN PUTTEN et al., 2012; RAEMAKERS et al., 2005; ZHAO et al., 2011) e melhoramento convencional (CEBALLOS et al., 2007; MORANTE et al., 2016). No entanto, o desenvolvimento de novas variedades ainda é limitado devido à dificuldade de florescimento dos genótipos elite, com conseqüente baixa produção de sementes e alta heterozigose, inerentes da cultura da mandioca, que faz com que procedimentos simples de transferência de características monogênicas, como é o caso do gene *waxy*, via retrocruzamentos, seja bastante complexa por resultar em alta recombinação genética e com isso não ser possível recuperar por completo o genoma do parental recorrente (CEBALLOS; KULAKOW; HERSHEY, 2012).

O caráter recessivo e o longo ciclo de melhoramento da cultura tornam onerosa a seleção de plantas para o fenótipo *waxy*, uma vez que no momento, a detecção precoce é realizada apenas via análises moleculares (AIEMNAKA et al., 2012). Assim,

metodologias rápidas que possibilitem a seleção indireta e a identificação da expressão do gene *waxy*, antes do seu armazenamento nas raízes da planta, podem otimizar o trabalho de seleção dos genótipos desejáveis e com isso permitir uma redução no ciclo de melhoramento da mandioca para desenvolvimento de novas variedades com esta característica.

No entanto, avaliações moleculares rápidas e automatizadas ainda possuem um custo relativamente elevado para a maioria dos programas de melhoramento genético. Por outro lado, a utilização de espectroscopia de infravermelho, surge como alternativa robusta e rápida para a identificação de genótipos contrastantes. A espectroscopia de infravermelho é uma técnica baseada na correlação entre as propriedades de vibração diferencial das moléculas (dependente da sua estrutura molecular) e suas implicações com a absorção de luz em determinadas frequências (NICOLAÏ et al., 2007). Embora essa técnica não seja capaz de detectar mudanças na molécula de DNA, mudanças sutis na expressão dos seus produtos gênicos podem ser detectadas, a exemplo da identificação de plantas transgênicas em soja que conferem resistência a herbicidas (LEE; CHOUNG, 2011) e variedades de tomate que possuem amadurecimento retardado (XIE et al., 2007).

Uma desvantagem inicial da espectroscopia no infravermelho próximo (NIR) é o alto custo de aquisição do equipamento, porém devido à multiplicidade de usos em materiais sólidos, pastosos e líquidos com alta versatilidade; elevada rapidez de análise; manuseio simples e prático sem necessidade de uso de reagentes complexos e alto custo, fazem com que o NIR seja um método bastante útil quando comparado com outros métodos químicos e analíticos. Além disso, como toda tecnologia, as possibilidades de uso do NIR vêm sendo aumentadas com o desenvolvimento de equipamentos de baixo custo que podem tornar as análises bem mais rápidas e práticas (KOSMOWSKI; WORKU, 2018).

Nas ciências agrárias, a espectrometria tem sido utilizada para classificação de solos (MOUAZEN et al., 2010), determinação de deficiência nutricional em plantas (BUTLER et al., 2017), identificação de espécies de plantas (CASTILLO et al., 2008; JIN et al., 2017; KOSMOWSKI; WORKU, 2018; LANG; ALMEIDA; COSTA, 2017)(CASTILLO et al., 2008; JIN et al., 2017; LANG; ALMEIDA; COSTA, 2017) discriminação de plantas transgênicas (LEE; CHOUNG, 2011; XIE et al., 2007) e determinação da origem geográfica de variedades de plantas (CHEN; ZHAO; LIN, 2009; LI et al., 2012). Para a realização destas análises é preciso reconhecer o padrão de espectros capaz de discriminar as classes que se pretende classificar/predizer. No

caso do NIR, os padrões de espectros são frequentemente detectados por métodos supervisionados de classificação. Variações de quadrados mínimos parciais (PLS - *partial least squares*) (BALABIN; SAFIEVA; LOMAKINA, 2011; MABOOD et al., 2017; TIXIER et al., 2017), máquinas de vetores de suporte (SVM - *support vector machine*) (BALABIN; SAFIEVA; LOMAKINA, 2011; GROMSKI et al., 2015; ZHENG; FU; YING, 2014), floresta aleatória (RF - *random forest*) (GROMSKI et al., 2015; LEE et al., 2013; STEVENS et al., 2013; XIN et al., 2012) e análise discriminante (ABDULRIDHA; EHSANI; DE CASTRO, 2016; BRITO et al., 2013; BUTLER et al., 2017; LANG; ALMEIDA; COSTA, 2017; LEE; CHOUNG, 2011; XIE et al., 2007) são algumas técnicas rotineiramente utilizados em espectroscopia para classificação e predição.

A espectrometria é uma técnica de fenotipagem que pode substituir avaliações laboriosas e/ou de maior custo muitas vezes com acurácia bastante compatível. Alguns exemplos de sucesso demonstram o potencial de uso da espectrometria, inclusive para identificação de sementes *waxy* em trigo (LAVINE; MIRJANKAR; DELWICHE, 2014). Inicialmente a identificação de sementes *waxy* em trigo era realizada com uso de técnicas químicas e moleculares, além da colorimetria. Entretanto, muitas vezes estas técnicas não conseguem mensurar estreitos teores de amilose e não são adequadas para a identificar linhagens parcialmente cerosas (à exceção da análise molecular). Por outro lado, o estabelecimento de modelos de predição via NIR permitiu a correta classificação em 100, 86 e 83 % de amostras *waxy*, parcial *waxyB1* e amostras não-*waxy* respectivamente, de maneira simples e rápida podendo substituir as avaliações convencionais com alta precisão (LAVINE; MIRJANKAR; DELWICHE, 2014).

Da mesma maneira em mandioca, algumas características utilizadas na seleção de plantas, sobretudo relacionadas a atributos das raízes, são de difícil mensuração e baixo rendimento de análise. Portanto, a utilização de técnicas para fenotipagem em larga escala, com precisão comparáveis as técnicas analíticas convencionais tem sido buscadas para uso nos programas de melhoramento genético (CABRERA-BOSQUET et al., 2012). Atualmente já encontra-se disponíveis versões portáteis do NIR, o que proporciona maior flexibilidade em realizar análises de matéria fresca de raízes a campo (IKEOGU et al., 2017). Nesse sentido, algumas aplicações de espectrometria foram recentemente utilizadas para conteúdo de matéria seca, carotenoides e ácido cianídrico, permitindo em uma única avaliação mensurar diferentes características das raízes (IKEOGU et al., 2017; SÁNCHEZ et al., 2014).

Este trabalho pretende associar espectros do NIR com componentes da folha de genótipos *waxy* e não-*waxy* de mandioca, para futuros trabalhos de classificação e seleção indireta de plantas desejáveis de forma precoce. O amido é basicamente constituído de dois tipos de polímeros, amilose e amilopectina, sintetizados nos cloroplastos como reserva de curto prazo (amido transitório) e nos amiloplastos como reserva de longo prazo (amido de armazenamento) (PFISTER; ZEEMAN, 2016). Nossa hipótese é que a amido transitório formado nos cloroplastos das folhas de mandioca estejam em concentração suficiente para análise por espectrometria e que diferenças na proporção amilose/amilopectina possam ser detectados. Portanto os objetivos deste trabalho são: *i*) avaliar o potencial da FT-NIR para a identificação de genótipos com amido *waxy*; e *ii*) selecionar e validar modelos de predição para classificação ideal para o fenótipo *waxy*.

MATERIAL E MÉTODOS

Material vegetal e preparo das amostras

Foram coletadas folhas completamente expandidas, sem sinais de doenças e ataque de insetos, de 134 genótipos para compor o conjunto de indivíduos não-*waxy* (variedades locais e melhoradas) e outros 162 genótipos para compor o conjunto de indivíduos *waxy* (população segregante). Todos os indivíduos fazem parte do Banco de Germoplasma de Mandioca (BAG-Mandioca) e do Programa de Melhoramento Genético da Embrapa Mandioca e Fruticultura (CNPMPF) (Material suplementar, Tabela S1 e S2). Os genótipos foram coletados em cinco campos experimentais da Embrapa Mandioca e Fruticultura e demais parceiros nos municípios de Laje e Cruz das Almas (Bahia, Brasil); Dourados (Mato Grosso do Sul, Brasil) e Londrina (Paraná, Brasil) (Tabela 1). Para as análises foram considerados 446 amostras de folhas (208 não-*waxy* e 238 *waxy*), referentes aos genótipos coletados em cada local (Tabela 1).

Para avaliação no NIR as folhas foram secas em estufa a 60° C e após este processo, foram maceradas e reduzidas a um pó fino com padronização do macerado em tecido tipo voil.

Análise espectrofotométrica e pré-processamentos dos dados

Os espectros foram mensurados por espectroscopia de infravermelho próximo com transformada de Fourier (FT-NIR) com uso do espectrômetro N500 FT-NIR® (Buchi). Os espectros foram coletados no modo de reflectância difusa a uma resolução

de 8 cm^{-1} , na faixa de número de ondas de $12.500 - 4000 \text{ cm}^{-1}$ ($1000 - 2500 \text{ nm}$). Uma calibração foi realizada automaticamente pelo equipamento a cada leitura de seis amostras.

Tabela 1: Relação dos locais de coleta e quantidade de genótipos avaliados de acordo com o fenótipo.

Local de coleta	Amido		Localização
	Não-waxy	Waxy	
Rio de Areia – Laje (Bahia)	94	60	13°08'47.6"S 39°17'58.0"W
Novo Horizonte – Laje (Bahia)	10	43	13°06'28.4"S 39°16'59.6"W
Cruz das Almas (Bahia)	88	-	12°40'47.7"S 39°05'20.9"W
Dourados (Mato Grosso do Sul)	16	102	22°16'36.0"S 54°48'56.0"W
Londrina (Paraná)	-	33	23°11'29.5"S 51°10'55.6"W
Total de amostras	208	238	

Foram realizadas duas leituras de cada amostra e após a obtenção dos espectros, os dados foram pré-processados pelo método da primeira derivada de Savitzky-Golay (SAVITZKY; GOLAY, 1964), variação normal padrão (SNV) e centralizados para remoção dos ruídos dos espectros e correção da dispersão dos dados. O pré-processamento foi realizado com auxílio do pacote *Prospectr* (STEVENS; RAMIREZ; LOPEZ, 2014) no ambiente de programação R v3.5 (R CORE TEAM, 2018).

Avaliação geral dos espectros

Após pré-processamento dos espectros a mediana das diferentes classes de amido e a análise de componentes principais de todos os indivíduos foi realizada para observar o potencial de agrupamento dos dados. A análise de componentes principais foi realizada no ambiente de programação R v3.5 (R CORE TEAM, 2018) com auxílio da função *prcomp*.

Técnicas de classificação

Para construção dos modelos foi utilizado o valor médio dos valores dos espectros obtidos para cada amostra. Adicionalmente a mediana para cada classe de amido (*waxy* e *não-waxy*) dos espectros foi realizada para definir faixas discrepantes. Desta forma duas avaliações foram realizadas: uma com toda faixa do espectro; e outra com apenas a faixa discrepante.

Cinco técnicas de reconhecimento de padrão supervisionados foram utilizadas para classificação dos dados: *Bayesian generalized linear model* – BGLM (GELMAN et al., 2008), *High dimensional discriminant analysis* – HDDA (BOUYEYRON; GIRARD; SCHMID, 2007), *Partial least squares* – PLS (WOLD, 1985), *Parallel random forest* – PRANDF (BREIMAN, 2001), e *Support vector machines with linear kernel* – SVM (CORTES; VAPNIK, 1995). Os algoritmos de classificação foram implementados no ambiente de programação R v3.5 (R CORE TEAM, 2018) com auxílio do pacote *Caret* (<http://caret.r-forge.r-project.org/>).

Validação externa e cruzada

Os dados pré-processados foram divididos em um conjunto para validação cruzada composto por 75% dados e o conjunto de validação externa com os 25% restantes; ambos com distribuição equitativa dos genótipos de acordo com o local de coleta e no tipo de amido e, portanto, representativos da diversidade dos ambientes analisados.

A validação cruzada foi realizada para obter uma estimativa mais robusta dos modelos de classificação, em 5 repetições com 10-folds cada. Portanto, 90% das amostras foram usadas como população de treinamento e 10% como população de validação. A seleção dos parâmetros finais dos modelos HDDA, PLS e RF foram baseadas nos maiores valores de acurácia. Para a análise discriminante HDDA foram testados *threshold* de 0.050, 0.175 e 0.300, enquanto que para o modelo PLS foram testados de um a três componentes; e no modelo PRANDF o número de preditores amostrados para divisão em cada nó foram 2, 54 e 1471.

A eficácia geral dos modelos de classificação foi avaliada com base nos valores médios da acurácia, pela estatística Kappa de Cohen (não ponderada) (COHEN, 1960) e pela análise ROC (*Receiver operating characteristics*). A acurácia foi obtida por: $\frac{vp+vn}{vp+fn+fp+vn}$ onde *vp* corresponde ao número de exemplos de classe corretamente reconhecidos (verdadeiros positivos), *vn* o número de exemplos corretamente reconhecidos que não pertencem à classe (verdadeiros negativos); *fp* exemplos que foram incorretamente atribuídos à classe (falsos positivos) e *fn* exemplos que não foram reconhecidos como exemplos de classe (falsos negativos). O índice de Kappa foi baseado no número de respostas concordantes, sendo definido por: $\frac{p_o+p_e}{1-p_e}$ onde p_o é a proporção de unidades concordantes; e p_e é a proporção de unidades para as quais o acordo é esperado por acaso. O resultado de Kappa pode ser interpretado da

seguinte forma: valores ≤ 0 indicam concordância insignificante; 0.01–0.20 fraca; 0.21–0.40 razoável; 0.41–0.60 moderada; 0.61–0.80 forte, e 0.81–1.00 quase perfeita (MCHUGH, 2012). ROC é obtida pela relação entre a sensibilidade em função da especificidade do modelo. A sensibilidade mede a probabilidade do classificador em acertar os verdadeiros positivos: $\frac{vp}{vp+fn}$, enquanto a especificidade mede a probabilidade em acertar os verdadeiros negativos: $\frac{vn}{fn+vn}$.

Os modelos de classificação foram avaliados aos pares pelo teste *t* em relação as diferenças de desempenho dos modelos com base em ROC. Os testes de significância foram corrigidos usando a correção de Bonferroni.

O conjunto de amostras para validação externa foi composto por 46 indivíduos não-*waxy* e 56 indivíduos *waxy*. Para a predição, o tipo de amido *waxy* foi classificado como classe positiva. O desempenho da predição foi avaliado com as medidas geradas a partir da matriz de confusão. Para tanto foram utilizadas a acurácia balanceada que minimiza o efeito de dados desbalanceados: $\frac{Sensibilidade+Especificidade}{2}$, a taxa de detecção: $\frac{vp+fn}{vp+fp+fn+vn}$ que corresponde a taxa de eventos verdadeiros, a prevalência de detecção: $\frac{vp+fp}{vp+fp+fn+vn}$ que corresponde a taxa de eventos classificados como verdadeiros e a precisão que é a relação de exemplos classificados como positivos corretamente $\frac{vp}{vp+fp}$.

RESULTADOS

Avaliação geral dos espectros

A análise de componentes principais (ACP) foi realizada para obter indicações preliminares sobre o possível agrupamento das amostras em relação aos tipos de amido ou ainda indicar componentes que pudessem caracterizar os genótipos em relação aos diferentes locais de coleta (Figura 1). Embora os dois primeiros componentes principais tenham sido responsáveis por grande parte da variação do espectro FT-NIR (71,93%), nenhum padrão claro de agrupamento dos genótipos de mandioca com base no tipo de amido foi observado. Também não foi possível encontrar nenhum padrão de agrupamento com base nos locais de coleta dos genótipos. A representação dos componentes PC1 e PC3 (53,04%) resultou em uma

melhor distinção tanto entre os genótipos (*waxy* e não-*waxy*) quanto entre os diferentes locais de coleta, mas ainda com muitas sobreposições (Figura 1).

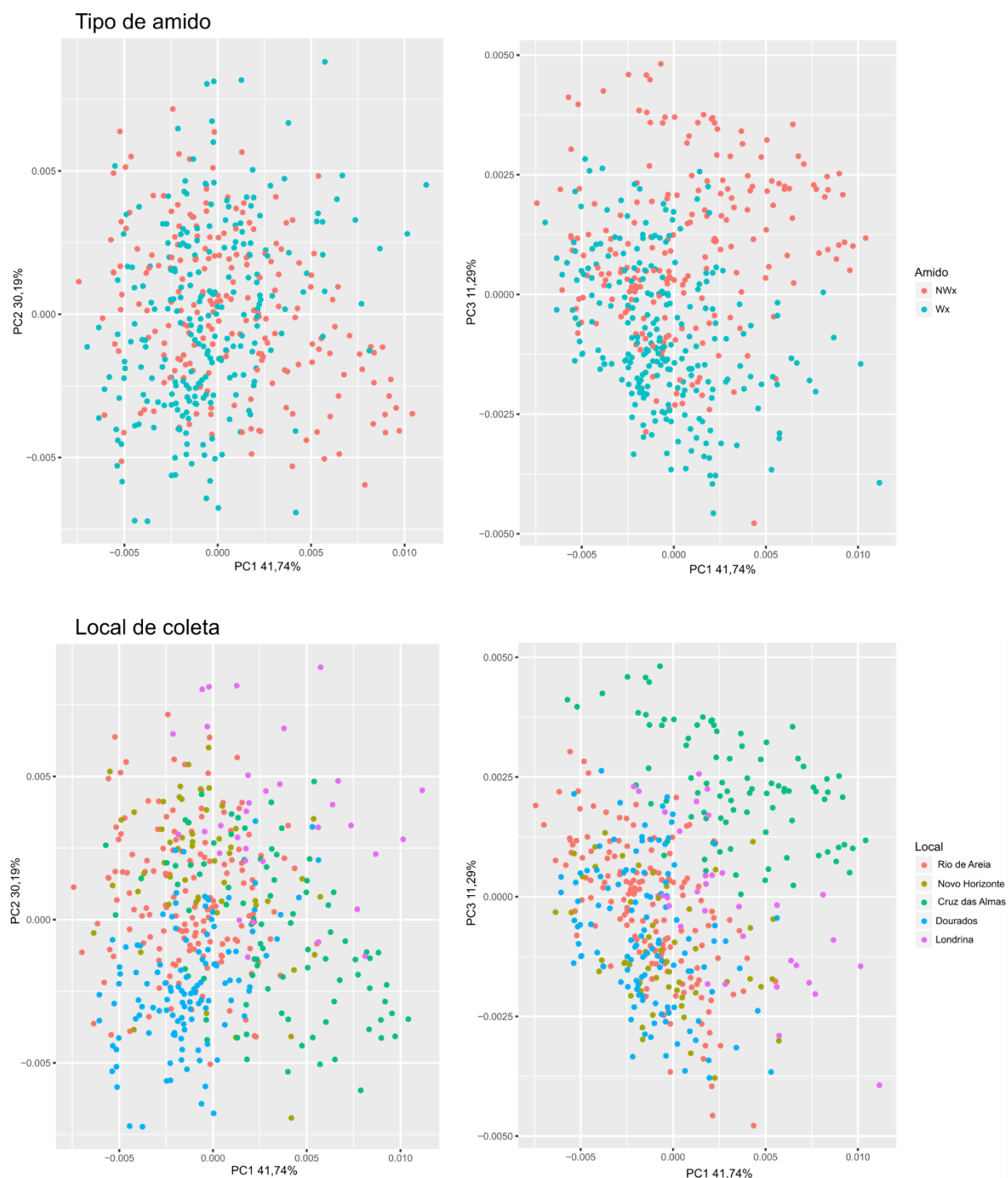


Figura 1. Análise de componentes principais com base nos espectros de infravermelho próximo com transformada de Fourier (FT-NIR) considerando amostras contrastantes para tipo de amido *waxy* (Wx) e não-*waxy* (NWx), bem como o local de coleta dos genótipos de mandioca.

O perfil da mediana dos espectros de absorvância, após pré-processamentos dos dados, foi bastante semelhante para as diferentes classes de genótipos (Wx – amido *waxy*; NWx – amido não-*waxy*) com principais regiões diferenciais dos espectros em destaque (Figura 2). No entanto, faz se necessário a utilização de técnicas quimiométricas destes dados para classificação dos genótipos *waxy* e não-*waxy*, uma vez que a simples diferenciação dos espectros pela mediana não diferencia completamente as duas classes de amido.

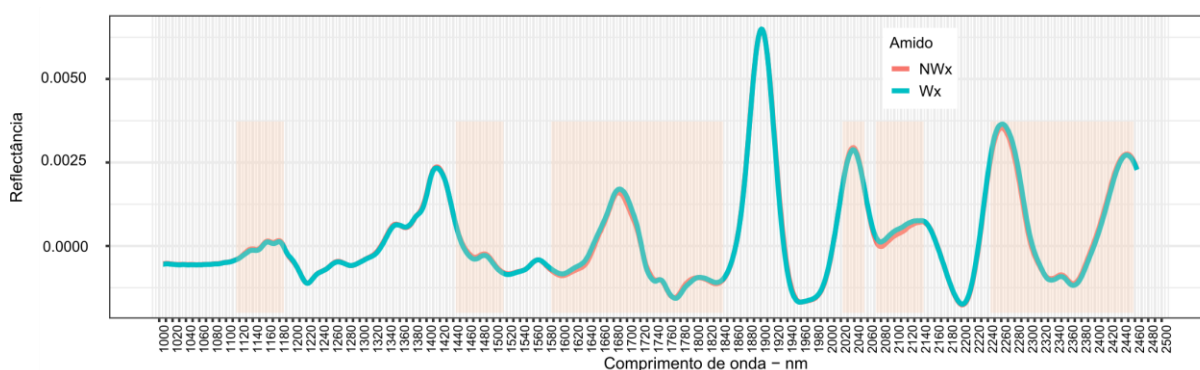


Figura 2. Distribuição da mediana dos espectros em amostras contrastantes para tipo de amido *waxy* (Wx) e não-*waxy* (NWx) obtidos por espectroscopia de infravermelho próximo com transformada de Fourier (FT-NIR). Em destaque as principais regiões do espectro contrastantes para tipo de amido.

Eficiência da classificação com base no tipo de amido de mandioca

O poder de discriminação dos modelos na validação cruzada foi avaliado a partir do número de amostras corretamente classificadas (acurácia) e a concordância da classificação (índice Kappa). Os modelos BGLM e SVM resultaram em alta capacidade de classificação dos genótipos *waxy* e não-*waxy*, com base nos altos valores de acurácia (0,94 e 0,92) e concordância Kappa quase perfeita (0,88 e 0,83) (Figura 3 A). Apesar dos elevados valores de acurácia dos modelos PRANDF, HDDA e PLS (0,77, 0,72 e 0,72 respectivamente) os valores médios de concordância foram de magnitudes mediana (0,55, 0,43 e 0,42 respectivamente).

O ranqueamento dos modelos de classificação com base apenas nos espectros com maior discrepância entre as classes foi bastante similar ao modelo completo, pois os modelos BGLM e SVM foram os mais acurados (Figura 3 B). No entanto, houve um pequeno decréscimo nos valores médios para acurácia de 0,86 e 0,85 e concordância Kappa de 0,71 e 0,80. Em contrapartida os modelos PRANDF, HDDA e PLS apresentaram maiores acurácia (0,83, 0,77 e 0,75) e forte (0,66) a moderada

(0,53 e 0,50) concordância Kappa, respectivamente, quando apenas as faixas espectrais selecionadas foram utilizadas nos modelos de classificação.

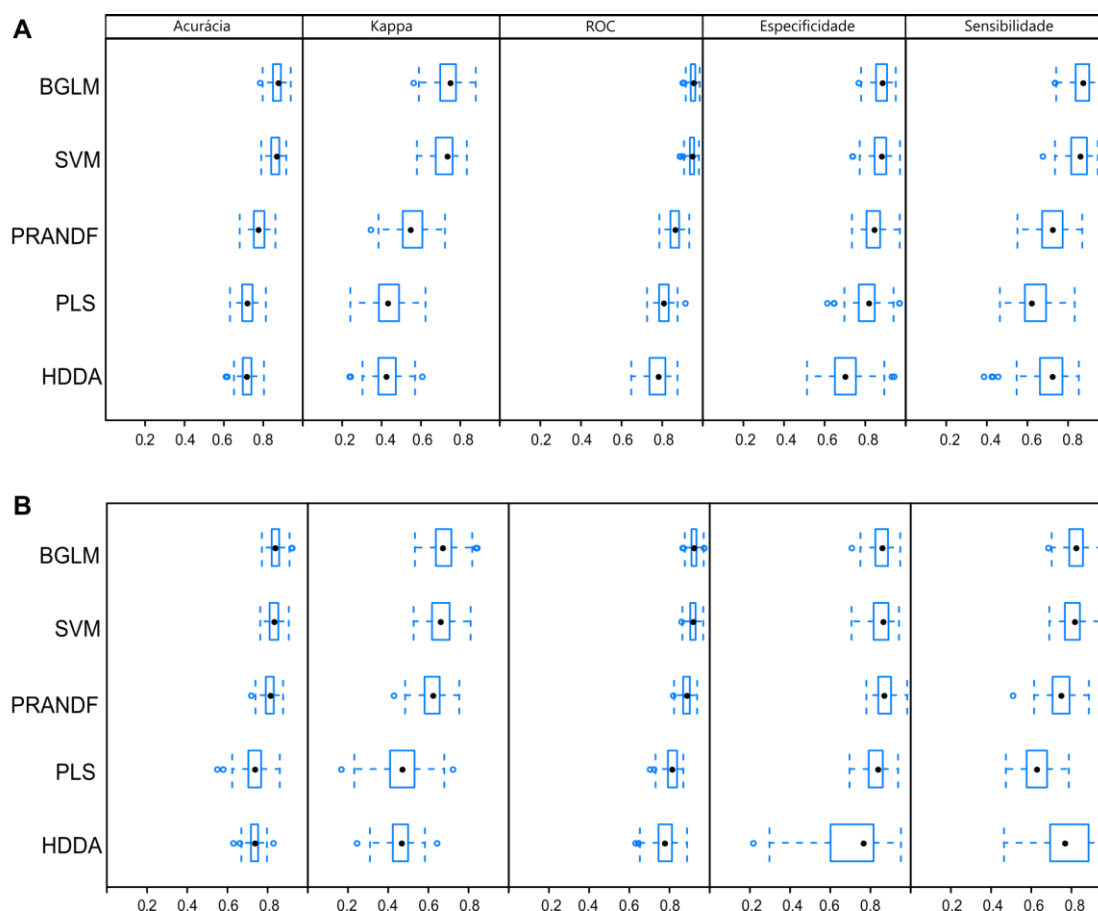


Figura 3. Avaliação da eficácia geral dos modelos de classificação com base nos espectros de infravermelho próximo com transformada de Fourier (FT-NIR) avaliados em amostras contrastantes para amido waxy (Wx) e não-waxy (NWx). **A** – Avaliação de todos os espectros; **B** – Avaliação dos espectros mais discrepantes entre as amostras waxy e não-waxy. Modelos de classificação: *Bayesian generalized linear* – BGLM; *support vector machines with linear kernel* – SVM; *parallel random forest* – PRANDF; *high dimensional discriminant analysis* – HDDA; e *partial least squares* – PLS.

Os modelos também foram avaliados pela probabilidade em acertar corretamente a classificação dos indivíduos waxy (sensibilidade) e não-waxy (especificidade). Desta forma, os modelos BGLM e SVM, apresentam altos valores de sensibilidade (média de 0,87 e 0,85 respectivamente) e especificidade (média de 0,88), e, por conseguinte, também apresentaram maior ROC (média 0,95 e 0,94 respectivamente) (Figura 3 A) quando foi considerado todo o espectro. Mesmo

comportamento foi observado quando apenas as faixas selecionadas foram avaliadas com valores médios de sensibilidade, especificidade e ROC maiores para os modelos BGLM (0,86, 0,89 e 0,93) e SVM (0,84, 0,89 e 0,93) (Figura 3 B).

Apesar dos demais modelos de classificação apresentarem valores medianos para os parâmetros de especificidade, os valores de sensibilidade foram baixos indicando que a classificação nos modelos PRANDF, HDDA (ambos 0,71) e PLS (0,61) são tendenciosas para atribuir os indivíduos a uma das duas classes (Figura 3 A). No entanto, índices melhores de sensibilidade foram obtidos pelos modelos HDDA (0,88), PRANDF (0,79) e PLS (0,68) quando apenas faixas discrepantes foram utilizadas (Figura 3 B).

Em relação à comparação par-a-par dos resultados ROC dos diferentes modelos de classificação, verificou-se a presença de diferenças significativas ($p < 0,05$) entre a maioria dos modelos avaliados com exceção da comparação dos modelos BGLM e SVM, que praticamente apresentaram a mesma eficiência de classificação dos genótipos de mandioca quanto ao tipo de amido em todo espectro (Figura 4 A) e para as faixas selecionadas (Figura 4 B).

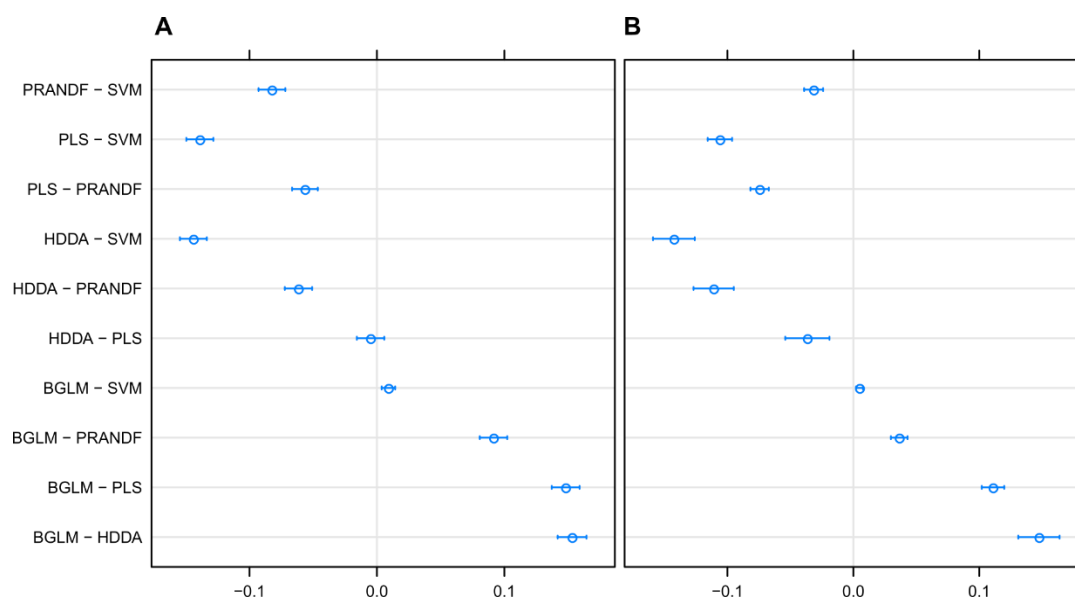


Figura 4. Diferenças significativa ($p < 0,05$) do parâmetro ROC (*Receiver operating characteristics*) entre os pares de modelos de classificação avaliados para diferenciação de genótipos para amido *waxy* (Wx) e não-*waxy* (NWx) de mandioca: *Bayesian generalized linear* – BGLM; *support vector machines with linear kernel* – SVM; *parallel random forest* – PRANDF; *high dimensional discriminant analysis* – HDDA; e *partial least squares* – PLS. **A** – Avaliação de todos os espectros; **B** – Avaliação dos espectros mais discrepantes entre as amostras *waxy* e não-*waxy*.

Validação externa dos modelos de classificação

A capacidade preditiva dos modelos foi avaliada com base nas medidas geradas a partir da matriz de confusão por validação externa, ou seja, a partir de amostras não utilizadas no treinamento: 446 genótipos foram utilizados para a construção dos modelos de classificação e 102 (62 *waxy* e 40 não-*waxy*) foram separados para a validação externa. Os modelos BGLM e SVM apresentaram o mesmo poder de discriminação dos genótipos em relação ao tipo de amido, enquanto o modelo PLS apresentou o pior desempenho nesta classificação (Tabela 2).

Tabela 2. Parâmetros de eficiência de diferentes modelos de classificação com base em todos os espectros de NIR avaliados em genótipos de mandioca contrastantes para amido *waxy* (Wx) e não-*waxy* (NWx), bem como a matriz de confusão avaliada em amostras externas de validação.

Parâmetros	BGLM ¹		SVM		PRANF		HDDA		PLS	
Acurácia balanceada	0,88		0,88		0,81		0,75		0,69	
Índice Kappa	0,76		0,76		0,62		0,50		0,39	
Especificidade	0,80		0,80		0,78		0,67		0,67	
Sensibilidade	0,95		0,95		0,84		0,82		0,71	
Precisão	0,85		0,85		0,82		0,75		0,73	
Prevalência de detecção	0,61		0,61		0,56		0,60		0,54	
Taxa de detecção	0,52		0,52		0,46		0,45		0,39	
Matriz de confusão	NWx	Wx	NWx	Wx	NWx	Wx	NWx	Wx	NWx	Wx
NWx	37	03	37	03	36	09	31	10	31	16
Wx	09	53	09	53	10	47	15	46	15	40

¹Modelos de classificação: *Bayesian generalized linear* – BGLM; *support vector machines with linear kernel* – SVM; *parallel random forest* – PRANF; *high dimensional discriminant analysis* – HDDA; e *partial least squares* – PLS.

Todos os modelos apresentaram alta taxa de amostras classificadas corretamente como fenótipo *waxy* (precisão variando de 0,85 nos modelos BGLM e SVM a 0,73 no modelo PLS). No entanto, a taxa de eventos verdadeiros e dos classificados como verdadeiro foi de menor magnitude, pois taxa de detecção variou de 0,52 no modelo BGLM e SVM a 0,39 no modelo PLS, enquanto a prevalência de detecção variou de 0,61 no modelo BGLM e SVM a 0,54 no modelo PLS. De modo geral, os modelos HDDA e PLS apresentaram menor capacidade de classificação correta dos genótipos de mandioca com base nos valores de especificidade (0,62) e índice Kappa (variando de 0,50 a 0,39).

Os modelos que apresentaram melhores valores relacionados aos parâmetros de classificação na validação cruzada (BGLM e SVM) também apresentaram melhores índices na validação externa (Tabela 2).

DISCUSSÃO

Desempenho dos modelos de classificação

Os modelos de classificação BGLM e SVM foram os mais eficientes em discriminar os diferentes genótipos de mandioca com tipos diferentes de amido pela análise indireta das folhas via NIR, em função dos maiores valores de acurácia, índice Kappa e ROC independente da faixa espectral utilizada. De modo geral, a precisão de um modelo é altamente dependente do conjunto de dados em análise e, portanto, é necessária a avaliação de vários modelos de classificação de forma a explorar diferenças nos algoritmos que possam traduzir em maior acurácia nos parâmetros de classificação. A descoberta de modelos que permitem alta acurácia seletiva é especialmente importante no presente trabalho, pois com a predição precoce dos genótipos de mandioca com amido *waxy*, apenas os clones com esta característica seriam selecionados em populações segregantes e mantidos no campo para colheitas e avaliações posteriores. Entretanto, se a acurácia de seleção for baixa, a possibilidade de descarte de plantas com a característica desejada é alta e vice-versa, e neste caso a seleção fenotípica convencional com base no teste de iodo nas raízes no momento da colheita (geralmente 12 meses após o plantio) seria mais vantajosa. Como, no presente estudo, a acurácia dos modelos BGLM e SVM foi bastante elevada (0,94 e 0,92, respectivamente), o uso destes modelos de classificação com base na análise dos espectros do NIR, possui um grande potencial para aplicação prática nos programas de melhoramento.

Outros trabalhos envolvendo espectrometria também apontaram que, modelos como SVM possuem alto poder de classificação e predição em comparação com outros algoritmos. De fato, alguns autores relataram melhor capacidade de distinção de genótipos de chá verde com base na sua origem geográfica (CHEN; ZHAO; LIN, 2009), e na detecção de adulterações em pasta de frutas (ZHENG; FU; YING, 2014). Nestes dois estudos, a acurácia de classificação do modelo SVM variou de 1,00 a 0,96 em comparação com a variação da acurácia de 0,67 a 0,98 nos modelos de classificação baseados em redes neurais e análises discriminantes. Em outro trabalho, o SVM apresentou maiores acurácias (>96%) em comparação com o

algoritmo *partial least squares–discriminant analysis* (PLS-DA) na classificação de variedades de milho (YANG et al., 2015). Por outro lado, o modelo de classificação *extreme learning machine* (EML) apresentou maior valor de acurácia na classificação de diferentes espécies de café, tipos de carne e origens de azeite (ZHENG; FU; YING, 2014), quando comparados ao SVM. Estes resultados mais uma vez demonstram a necessidade de validar modelos de classificação para cada tipo de dado. Além disso, até o presente, não foram encontrados trabalhos na literatura referentes ao uso do modelo BGLM na análise de dados de espectrometria e, portanto, nossos resultados indicam que este modelo de classificação também possui alto potencial para a classificação de dados dessa natureza.

Apesar do baixo desempenho dos modelos PLS na classificação de genótipos *waxy* e *não-waxy*, em outras abordagens da espectrometria, o PLS apresentou excelentes predições, por exemplo, na identificação de espécies com material fibroso ($r^2 = 0,98$) (TIXIER et al., 2017). De forma semelhante, outros modelos baseados em *linear discriminant analysis* (LDA) têm sido utilizadas para distinguir famílias, gêneros, e espécies de plantas arbóreas, com elevada acurácia na identificação de gêneros e famílias (0,91 e 0,95, respectivamente) (LANG; ALMEIDA; COSTA, 2017). Modelos baseado em *partial least-squares discriminant analysis* (PLSDA) também tem apresentado grande poder de discriminação de transgênicos em tomate (XIE et al., 2007) e soja (LEE; CHOUNG, 2011) com acurácia de 1,00 e 0,97, respectivamente.

Embora um dos objetivos do trabalho tenha sido selecionar modelos com capacidade de discriminação entre os diferentes tipos de amido, os modelos SVM e BGLM apresentaram resultados muito equivalentes, mesmo utilizando algoritmos de classificação completamente diferentes. O modelo SVM escolhe o hiperplano que melhor representa a separação entre as duas classes para classificação os dados. Já o BGLM ajusta modelos lineares generalizados com a distribuição prévia de Student, incorporando um algoritmo de maximização de expectativas (EM) aproximado aos mínimos quadrados ponderados iterativamente (GELMAN et al., 2008). Apesar das semelhanças nos parâmetros de qualidade das classificações, o modelo SVM apresentou menor tempo de processamento com uma precisão comparável ao BGLM sendo, portanto, melhor indicado para avaliações futuras do tipo de amido em mandioca.

No presente trabalho utilizou-se um conjunto de dados para validação externa, tendo como objetivo verificar se o modelo de classificação previamente construído com outro conjunto de dados poderia discriminar corretamente um conjunto de novos

indivíduos sem atribuição de classe. Como resultado, os algoritmos de treinamento SVM e BGLM construíram modelos que atribuíram corretamente 85% dos indivíduos *waxy*. Os demais modelos apresentaram classificação bastante inferior (<75% para os modelos PRANDF e HDDA, e apenas <65% para o modelo PLS). Os modelos BGLM e SVM apresentaram menor taxa de falsos negativos ou “erro tipo II”, sendo, portanto, mais indicados no processo seletivo, pois é preferível cometer erros na seleção de indivíduos não-*waxy* do que no descarte de indivíduos com amido *waxy*.

Uso da espectrometria FT-NIR em tecidos foliares

A classificação de indivíduos utilizando a espectrometria FT-NIR para captar de forma indireta diferenças na constituição química das plantas como resultado da expressão de diferentes genes tem sido utilizada nas mais diversas áreas das ciências biológicas e agrárias. A utilização da folha é comum uma vez que sua estrutura, fisiologia, bioquímica e componentes da parede celular estão relacionados a espécie e as condições do ambiente onde a planta está inserida. Folhas secas e maceradas já foram utilizadas na distinção de espécies do gênero *Eucalyptus* (CASTILLO et al., 2008), na distinção de origem geográfica de chá verde (CHEN; ZHAO; LIN, 2009) e na distinção de variedades de *Camellia sinensis* L. (CHEN et al., 2008). Esses autores relataram excelentes acurácias de classificação, de 1,00 na distinção de espécies de eucalipto e identificação da origem de chá e 0,94 na identificação de variedades de chá verde.

De acordo com Lang; Almeida; Costa (2017) a identidade taxonômica de plantas, pode ser discriminada entre diversas famílias, gêneros e espécies de árvores da Amazônia a partir de ramos e folhas secas digitalizados pelo FT-NIR sem maceração prévia (LANG; ALMEIDA; COSTA, 2017). Como amostra de análise, uma porção dos ramos e três folhas secas digitalizados por espécime, resultaram em acurácia de 0,91 e 0,94, respectivamente. Esta mesma estratégia foi utilizada para distinção entre os gêneros *Eschweilera* e *Corythophora*, cuja porcentagem de identificação corretas das espécies foi de 0,99 (DURGANTE et al., 2013), bem como na predição de três espécies de miscanto com acurácia completa quando se utilizou o modelo *least squares support vector machine regression* (LSSVR) (JIN et al., 2017).

Outras partes da planta também podem ser utilizadas para classificação e predição de determinados fenótipos de interesse. Como exemplo, alguns autores tem utilizado sementes de cevada, grão-de-bico e sorgo para distinguir cultivares, com excelentes acurácias (>0,87) (KOSMOWSKI; WORKU, 2018). Sementes de plantas

transgênicos também já foram discriminadas de plantas convencionais com uso da espectrometria com 0,97 de acurácia pelo modelo PLSDA (LEE; CHOUNG, 2011). Portanto, estes trabalhos evidenciam, a eficiência da espectrometria FT-NIR em diferenciar indivíduos com características específicas mesmo havendo apenas diferenças sutis na constituição dos tecidos avaliados.

Perspectivas de uso do NIR na classificação de genótipos de mandioca com amido waxy

O presente estudo baseia-se na hipótese de que é possível utilizar o NIR para diferenciar genótipos de mandioca com características diferenciadas de amido, sobretudo da relação amilose/amilopectina, uma vez que cada tipo de ligação química tem vibração única e portanto pode ser mensurada indiretamente com base nos espectros de infravermelho (PASQUINI, 2003). Os resultados indicaram dois modelos de classificação com grande aplicabilidade na discriminação do fenótipo *waxy* do não-*waxy* de maneira rápida e com baixo custo.

A espectrometria tem demonstrado alto potencial de discriminação do local de origem de plantas como *Camellia sinensis* e *Codonopsis pilosula* (CHEN; ZHAO; LIN, 2009; LI et al., 2012) e de modo geral, a inclusão de dados oriundos de diferentes ambientes melhora a robustez do modelo de predição (DIAGO et al., 2018). Como a mandioca é cultivada em diferentes condições edafoclimáticas, os modelos de classificação e predição devem ser acurados o suficiente para identificar os indivíduos *waxy* considerando estas diferentes condições ambientais. No presente estudo, foram utilizadas amostras de diferentes locais com o objetivo de captar estes efeitos ambientais, tornando a predição pelos modelos mais acurada. Entretanto, não foi observada distinção clara entre os locais de coleta por meio da PCA, possivelmente como um resultado do desbalanceamento do número de indivíduos por local. Embora este não tenha sido um dos objetivos deste trabalho, recomenda-se que a análise futura de um grande número de ambientes seja utilizada para distinguir a origem dos genótipos de mandioca, de forma a compreender melhor os efeitos ambientais nos modelos de classificação dos fenótipos *waxy* e não-*waxy*.

De modo geral a análise de folhas secas sem maceração, via NIR, tiveram como objetivo desenvolver modelos para classificação/predição em materiais vegetais em condições próximas as de material depositados em herbário (DURGANTE et al., 2013; LANG; ALMEIDA; COSTA, 2017). A escolha das folhas de mandioca secas e maceradas deveu-se à facilidade de armazenamento e manuseio destas amostras

foliares após a coleta. Neste caso, os genótipos tiveram que ser plantados em campo para posterior análise no NIR. Contudo, no futuro, será necessário avaliar a acurácia de classificação de genótipos *waxy* e *não-waxy* após a obtenção dos cruzamentos desejados, tendo em vista que o interesse dos melhoristas é gerar grandes populações segregantes (com milhares de indivíduos) e proceder o *screening* dos genótipos de interesse antes do plantio no campo. Isto poderia otimizar o espaço destinado à avaliação apenas dos genótipos com a característica desejada, e com isso reduzir o uso de insumos, agroquímicos e manejo necessário para o desenvolvimento das plantas. A viabilidade desta estratégia baseia nos resultados promissores em outras espécies vegetais, a exemplo das análises não-destrutiva de sementes de milho que tiveram como foco a avaliação da viabilidade das sementes (AMBROSE et al., 2016), identificação de tipos *waxy* e *não-waxy* em trigo (LAVINE; MIRJANKAR; DELWICHE, 2014), monitoramento de parâmetros de qualidade de diferentes variedades de oliveira (FERNÁNDEZ-ESPINOSA, 2016) e classificação de variedades de videiras (GUTIÉRREZ et al., 2015).

Independentemente do momento e tecido analisado, a metodologia FT-NIR demonstrou alto potencial de aplicabilidade, sendo uma ferramenta rápida e eficaz quanto à identificação de genótipos *waxy*, para uso no programa de melhoramento da mandioca. O desenvolvimento de metodologias rápidas para a identificação de indivíduos contrastantes para o fenótipo *waxy* por meio de algoritmos de classificação não supervisionados com base em FT-NIR pode trazer grandes benefícios para o desenvolvimento de novas variedades de mandioca com esta característica. No contexto da seleção precoce apenas os clones com esta característica poderão ser selecionados em populações segregantes ainda em casa de vegetação, reduzindo custos de implantação e manutenção no campo. Da mesma maneira para populações segregantes já estabelecidas, a seleção pode ocorrer antes do armazenamento das raízes, diminuindo o tempo na seleção de indivíduos para o próximo ciclo.

REFERÊNCIAS

ABDULRIDHA, J.; EHSANI, R.; DE CASTRO, A. Detection and Differentiation between Laurel Wilt Disease, Phytophthora Disease, and Salinity Damage Using a Hyperspectral Sensing Technique. **Agriculture**, v. 6, n. 56, p. 1–13, 2016.

AIEMNAKA, P.; WONGKAEW, A.; CHANTHAWORN, J.; NAGASHIMA, S. K.;

BOONMA, S.; AUTHAPUN, J.; JENWEERAWAT, S.; KONGSILA, P.; KITTIPADAKUL, P.; NAKASATHIEN, S.; SREEWONGCHAI, T.; WANNARAT, W.; VICHUKIT, V.; LÓPEZ-LAVALLE, L. A. B.; CEBALLOS, H.; ROJANARIDPICHED, C.; PHUMICHA, C. Molecular characterization of a spontaneous waxy starch mutation in cassava. **Crop Science**, v. 52, n. 5, p. 2121–2130, 2012.

AMBROSE, A.; LOHUMI, S.; LEE, W. H.; CHO, B. K. Comparative nondestructive measurement of corn seed viability using Fourier transform near-infrared (FT-NIR) and Raman spectroscopy. **Sensors and Actuators, B: Chemical**, v. 224, p. 500–506, 2016.

BALABIN, R. M.; SAFIEVA, R. Z.; LOMAKINA, E. I. Near-infrared (NIR) spectroscopy for motor oil classification: From discriminant analysis to support vector machines. **Microchemical Journal**, v. 98, n. 1, p. 121–128, 2011.

BOUYEYRON, C.; GIRARD, S.; SCHMID, C. High-dimensional discriminant analysis. **Communications in Statistics - Theory and Methods**, v. 36, n. 14, p. 2607–2623, 2007.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

BREUNINGER, W. F.; PIYACHOMKWAN, K.; SRIROTH, K. Tapioca/Cassava Starch. In: **Starch**. [s.l.] : Elsevier, 2009. p. 541–568.

BRITO, A. L. B.; BRITO, L. R.; HONORATO, F. A.; PONTES, M. J. C.; PONTES, L. F. B. L. Classification of cereal bars using near infrared spectroscopy and linear discriminant analysis. **Food Research International**, 2013.

BUTLER, H. J.; ADAMS, S.; MCAINSH, M. R.; MARTIN, F. L. Detecting nutrient deficiency in plant systems using synchrotron Fourier-transform infrared microspectroscopy. **Vibrational Spectroscopy**, v. 90, p. 46–55, 2017.

CABRERA-BOSQUET, L.; CROSSA, J.; VON ZITZEWITZ, J.; SERRET, M. D.; LUIS ARAUS, J. High-throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding Converge. **Journal of Integrative Plant Biology**, v. 54, n. 5, p. 312–

320, 2012.

CASTILLO, R.; CONTRERAS, D.; FREER, J.; RUIZ, J.; VALENZUELA, S. Supervised pattern recognition techniques for classification of Eucalyptus species from leaves NIR spectra. **Journal of the Chilean Chemical Society**, v. 53, n. 4, p. 1709–1713, 2008.

CEBALLOS, H.; KULAKOW, P.; HERSHEY, C. Cassava Breeding: Current Status, Bottlenecks and the Potential of Biotechnology Tools. **Tropical Plant Biology**, v. 5, n. 1, p. 73–87, 2012.

CEBALLOS, H.; SÁNCHEZ, T.; MORANTE, N.; FREGENE, M.; DUFOUR, D.; SMITH, A. M.; DENYER, K.; PÉREZ, J. C.; CALLE, F.; MESTRES, C. Discovery of an amylose-free starch mutant in cassava (*Manihot esculenta* Crantz). **Journal of Agricultural and Food Chemistry**, v. 55, n. 18, p. 7469–7476, 2007.

CHEN, Q.; ZHAO, J.; LIN, H. Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 72, n. 4, p. 845–850, 2009.

CHEN, Q.; ZHAO, J.; LIU, M.; CAI, J. Nondestructive Identification of Tea (*Camellia sinensis* L.) varieties using FT-NIR spectroscopy and pattern recognition. **Czech Journal of Food Sciences**, v. 26, n. 5, p. 360–367, 2008.

COHEN, J. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37–46, 1960.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, v. 20, n. 3, p. 273–297, 1995.

DIAGO, M. P.; FERNÁNDEZ-NOVALES, J.; GUTIÉRREZ, S.; MARAÑÓN, M.; TARDAGUILA, J. Development and Validation of a new methodology to assess the vineyard water status by on-the-go Near infrared spectroscopy. **Frontiers in Plant Science**, v. 9, p. 1–13, 2018.

DURGANTE, F. M.; HIGUCHI, N.; ALMEIDA, A.; VICENTINI, A. Species Spectral Signature: Discriminating closely related plant species in the Amazon with Near-Infrared Leaf-Spectroscopy. **Forest Ecology and Management**, v. 291, p. 240–248, 2013.

FERNÁNDEZ-ESPINOSA, A. J. Combining PLS regression with portable NIR spectroscopy to on-line monitor quality parameters in intact olives for determining optimal harvesting time. **Talanta**, v. 148, p. 216–228, 2016.

GELMAN, A.; JAKULIN, A.; PITTAU, M. G.; SU, Y. S. A weakly informative default prior distribution for logistic and other regression models. **Annals of Applied Statistics**, v. 2, n. 4, p. 1360–1383, 2008.

GROMSKI, P. S.; XU, Y.; HOLLYWOOD, K. A.; TURNER, M. L.; GOODACRE, R. The influence of scaling metabolomics data on model classification accuracy. **Metabolomics**, v. 11, n. 3, p. 684–695, 2015.

GUTIÉRREZ, S.; TARDAGUILA, J.; FERNÁNDEZ-NOVALES, J.; DIAGO, M. P. Support vector machine and artificial neural network models for the classification of grapevine varieties using a portable NIR spectrophotometer. **PLoS ONE**, v. 10, n. 11, p. 1–15, 2015.

IKEOGU, U. N.; DAVRIEUX, F.; DUFOUR, D.; CEBALLOS, H.; EGESI, C. N.; JANNINK, J. L. Rapid analyses of dry matter content and carotenoids in fresh cassava roots using a portable visible and near infrared spectrometer (Vis/NIRS). **PLoS ONE**, v. 12, n. 12, p. 1–17, 2017.

JIN, X.; CHEN, X.; XIAO, L.; SHI, C.; CHEN, L.; YU, B.; YI, Z.; YOO, J. H.; HEO, K.; YU, C. Y.; YAMADA, T.; SACKS, E. J.; PENG, J. Application of visible and near-infrared spectroscopy to classification of *Miscanthus* species. **PLoS ONE**, v. 12, n. 4, p. 1–12, 2017.

KARLSTRÖM, A.; CALLE, F.; SALAZAR, S.; MORANTE, N.; DUFOUR, D.; CEBALLOS, H. Biological implications in Cassava for the production of amylose-free starch: Impact on root yield and related traits. **Frontiers in Plant Science**, v. 7, p. 1–

8, 2016.

KOEHORST-VAN PUTTEN, H. J. J.; WOLTERS, A. M. a; PEREIRA-BERTRAM, I. M.; VAN DEN BERG, H. H. J.; VAN DER KROL, A. R.; VISSER, R. G. F. Cloning and characterization of a tuberous root-specific promoter from cassava (*Manihot esculenta* Crantz). **Planta**, v. 236, n. 6, p. 1955–1965, 2012.

KOSMOWSKI, F.; WORKU, T. Evaluation of a miniaturized NIR spectrometer for cultivar identification: The case of barley, chickpea and sorghum in Ethiopia. **PLoS ONE**, v. 13, n. 3, p. 1–17, 2018.

LANG, C.; ALMEIDA, D. R. A.; COSTA, F. R. C. Discrimination of taxonomic identity at species, genus and family levels using Fourier Transformed Near-Infrared Spectroscopy (FT-NIR). **Forest Ecology and Management**, v. 406, p. 219–227, 2017.

LAVINE, B. K.; MIRJANKAR, N.; DELWICHE, S. Classification of the waxy condition of durum wheat by near infrared reflectance spectroscopy using wavelets and a genetic algorithm. **Microchemical Journal**, v. 117, p. 178–182, 2014.

LEE, J. H.; CHOUNG, M.-G. Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. **Food Chemistry**, v. 126, n. 1, p. 368–373, 2011.

LEE, S.; CHOI, H.; CHA, K.; CHUNG, H. Random forest as a potential multivariate method for near-infrared (NIR) spectroscopic analysis of complex mixture samples: Gasoline and naphtha. **Microchemical Journal**, v. 110, p. 739–748, 2013.

LI, B.; WEI, Y.; DUAN, H.; XI, L.; WU, X. Discrimination of the geographical origin of *Codonopsis pilosula* using near infrared diffuse reflection spectroscopy coupled with random forests and k-nearest neighbor methods. **Vibrational Spectroscopy**, v. 62, p. 17–22, 2012.

MABOOD, F.; JABEEN, F.; HUSSAIN, J.; AL-HARRASI, A.; HAMAED, A.; AL MASHAYKHI, S. A. A.; AL RUBAIEY, Z. M. A.; MANZOOR, S.; KHAN, A.; HAQ, Q.

M. I.; GILANI, S. A.; KHAN, A. FT-NIRS coupled with chemometric methods as a rapid alternative tool for the detection & quantification of cow milk adulteration in camel milk samples. **Vibrational Spectroscopy**, v. 92, p. 245–250, 2017.

MCHUGH, M. L. Interrater reliability: the kappa statistic. **Biochemia Medica**, v. 22, n. 3, p. 276–282, 2012.

MOORTHY, S. N. Physicochemical and Functional Properties of Tropical Tuber Starches: A Review. **Starch - Stärke**, v. 54, n. 12, p. 559–592, 2002.

MORANTE, N.; CEBALLOS, H.; SÁNCHEZ, T.; ROLLAND-SABATÉ, A.; CALLE, F.; HERSHEY, C.; GIBERT, O.; DUFOUR, D. Discovery of new spontaneous sources of amylose-free cassava starch and analysis of their structure and techno-functional properties. **Food Hydrocolloids**, v. 56, p. 383–395, 2016.

MOUAZEN, A. M.; KUANG, B.; DE BAERDEMAEKER, J.; RAMON, H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. **Geoderma**, v. 158, n. 1–2, p. 23–31, 2010.

NICOLAÏ, B. M.; BEULLENS, K.; BOBELYN, E.; PEIRS, A.; SAEYS, W.; THERON, K. I.; LAMMERTYN, J. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. **Postharvest Biology and Technology**, v. 46, n. 2, p. 99–118, 2007.

OLOMO, O. V.; AJIBOLA, O. O. Processing factors affecting the yield and physicochemical properties of starches from cassava chips and flour. **AMA, Agricultural Mechanization in Asia, Africa and Latin America**, v. 37, n. 1, p. 65–69, 2006.

PASQUINI, C. Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198–219, 2003.

PFISTER, B.; ZEEMAN, S. C. Formation of starch in plant cells. **Cellular and**

Molecular Life Sciences, v. 73, n. 14, p. 2781–2807, 2016.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**, 2018. Disponível em: <<https://www.r-project.org/>>. Acesso em: março de 2018.

RAEMAKERS, K.; SCHREUDER, M.; SUURS, L.; FURRER-VERHORST, H.; VINCKEN, J.-P.; DE VETTEN, N.; JACOBSEN, E.; VISSER, R. G. F. Improved Cassava Starch by Antisense Inhibition of Granule-bound Starch Synthase I. **Molecular Breeding**, v. 16, n. 2, p. 163–172, 2005.

SÁNCHEZ, T.; CEBALLOS, H.; DUFOUR, D.; ORTIZ, D.; MORANTE, N.; CALLE, F.; ZUM FELDE, T.; 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.; DUFOUR, D.; MORENO, I. X.; CEBALLOS, H. Comparison of pasting and gel stabilities of waxy and normal starches from potato, maize, and rice with those of a novel waxy cassava starch under thermal, chemical, and mechanical stress. **Journal of Agricultural and Food Chemistry**, v. 58, n. 8, p. 5093–5099, 2010.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical chemistry**, v. 36, n. 8, p. 1627–1639, 1964.

STEVENS, A.; NOCITA, M.; TÓTH, G.; MONTANARELLA, L.; VAN WESEMAEL, B. Prediction of soil organic carbon at the European scale by visible and Near InfraRed reflectance Spectroscopy. **PLoS ONE**, v. 8, n. 6, p. e66409, 2013.

STEVENS, A.; RAMIREZ, L.; LOPEZ, –. **An introduction to the prospectr package**, 2014.

TIXIER, A.; HENRY, R.; LAI, C.; HAN, G.; VIA, B.; JIANG, W.; ZHOU, C.; SWAIN, T.; FAN, Z.; LIU, S. Classification and Identification of Plant Fibrous Material with Different Species Using near Infrared Technique—A New Way to Approach Determining Biomass Properties Accurately within Different Species. **Frontiers in**

Plant Science, v. 7, 2017.

WANG, S.; LI, C.; COPELAND, L.; NIU, Q.; WANG, S. Starch Retrogradation: A Comprehensive Review. **Comprehensive Reviews in Food Science and Food Safety**, v. 14, n. 5, p. 568–585, 2015.

WOLD, H. Partial least squares. **Encyclopedia of Statistical Sciences**, n. 2, p. 581–591, 1985.

XIE, L.; YING, Y.; YING, T.; YU, H.; FU, X. Discrimination of transgenic tomatoes based on visible/near-infrared spectra. **Analytica Chimica Acta**, v. 584, n. 2, p. 379–384, 2007.

XIN, N.; GU, X. F.; WU, H.; HU, Y. Z.; YANG, Z. L. Discrimination of raw and processed *Dipsacus asperoides* by near infrared spectroscopy combined with least squares-support vector machine and random forests. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 89, p. 18–24, 2012.

ZHAO, S.-S.; DUFOUR, D.; SÁNCHEZ, T.; CEBALLOS, H.; ZHANG, P. Development of waxy cassava with different Biological and physico-chemical characteristics of starches for industrial applications. **Biotechnology and Bioengineering**, v. 108, n. 8, p. 1925–1935, 2011.

ZHENG, W.; FU, X.; YING, Y. Spectroscopy-based food classification with extreme learning machine. **Chemometrics and Intelligent Laboratory Systems**, v. 139, p. 42–47, 2014.

MATERIAL SUPLEMENTAR

Tabela S1: Relação de genótipos de mandioca utilizados na análise de Espectroscopia de Infravermelho Próximo com Transformada de Fourier (FT-NIR), classificados de acordo com o tipo de amido (*waxy* – Wx e *nãowaxy* - NWx), bem como o local de coleta.

Acesso	Local	Tipo	Acesso	Local	Tipo	Acesso	Local	Tipo
4271	3	NWx	2014-019-006	1	NWx	BGM-0066	3	NWx
2011-024-015	4	NWx	2014-022-007	1	NWx	BGM-0067	3	NWx
2011-034-041	4	NWx	2014-023-020	1	NWx	BGM-0089	3	NWx
2011-034-045	4	NWx	2014-025-019	1	NWx	BGM-0115	3	NWx
2011-034-064	4	NWx	2014-025-042	1	NWx	BGM-0120	3	NWx
2011-034-069	4	NWx	2014-025-064	1	NWx	BGM-0122	3	NWx
2012-107-002	4	NWx	2014-026-006	1	NWx	BGM-0144	3	NWx
2014-001-024	1	NWx	2014-026-033	1	NWx	BGM-0148	3	NWx
2014-001-026	1	NWx	2014-026-035	1	NWx	BGM-0163	3	NWx
2014-003-005	1	NWx	2014-027-004	1	NWx	BGM-0184	3	NWx
2014-006-002	1	NWx	2014-034-011	1	NWx	BGM-0190	3	NWx
2014-006-018	1	NWx	2014-035-020	1	NWx	BGM-0955	3	NWx
2014-006-021	1	NWx	98150-06	1,2,4,3	NWx	BGM-0957	3	NWx
2014-006-032	1	NWx	Baianinha	4	NWx	BGM-0958	3	NWx
2014-006-055	1	NWx	BGM-0004	3	NWx	BGM-0959	3	NWx
2014-008-011	1	NWx	BGM-0014	3	NWx	BGM-0960	3	NWx
2014-008-024	1	NWx	BGM-0015	3	NWx	BGM-0962	3	NWx
2014-009-007	1	NWx	BGM-0019	3	NWx	BGM-0967	3	NWx
2014-009-008	1	NWx	BGM-0020	3	NWx	BGM-0968	3	NWx
2014-009-011	1	NWx	BGM-0021	3	NWx	BGM-0971	3	NWx
2014-010-002	1	NWx	BGM-0022	3	NWx	BGM-0973	3	NWx
2014-010-006	1	NWx	BGM-0022tr	3	NWx	BGM-0974	3	NWx
2014-010-011	1	NWx	BGM-0024	3	NWx	BGM-0975	3	NWx
2014-011-007	1	NWx	BGM-0028	3	NWx	BGM-0978	3	NWx
2014-011-011	1	NWx	BGM-0029	3	NWx	BGM-0980	3	NWx
2014-012-007	1	NWx	BGM-0030	3	NWx	BGM-0981	3	NWx
2014-012-013	1	NWx	BGM-0032	3	NWx	BGM-0982	3	NWx
2014-013-007	1	NWx	BGM-0035	3	NWx	BGM-0984	3	NWx
2014-013-024	1	NWx	BGM-0036	3	NWx	BGM-0985	3	NWx
2014-013-028	1	NWx	BGM-0037	3	NWx	BGM-0986	3	NWx
2014-013-030	1	NWx	BGM-0040	3	NWx	BGM-0989	3	NWx
2014-013-041	1	NWx	BGM-0042	3	NWx	BGM-0992	3	NWx
2014-013-050	1	NWx	BGM-0044	3	NWx	BGM-0994	3	NWx
2014-013-051	1	NWx	BGM-0046	3	NWx	BGM-0995	3	NWx
2014-015-009	1	NWx	BGM-0054	3	NWx	BGM-0996	3	NWx
2014-015-014	1	NWx	BGM-0056	3	NWx	BGM-0998	3	NWx
2014-018-027	1	NWx	BGM-0058	3	NWx	BGM-1006	3	NWx
2014-018-030	1	NWx	BGM-0061	3	NWx	BGM-1023	3	NWx
2014-018-039	1	NWx	BGM-0062	3	NWx	BGM-1024	3	NWx
2014-019-003	1	NWx	BGM-0063	3	NWx	BGM-1025	3	NWx

Tabela S1: Cont....

Acesso	Local	Tipo	Acesso	Local	Tipo	Acesso	Local	Tipo
BGM-1029	3	NWx	SP-0728-32	1	NWx	Wx-6519-03	5	Wx
BGM-1031	3	NWx	SP-0728-57	1	NWx	Wx-6519-04	4	Wx
BGM-2105	3	NWx	SP-1042-07	1	NWx	Wx-6526-01	1,2	Wx
BGM-2265	3	NWx	SP-1042-18	1	NWx	Wx-6618-01	5	Wx
BGM-2276	3	NWx	SP-1148-07	1	NWx	Wx-6668-03	4	Wx
BGM-2284	3	NWx	SP-1378-43	1	NWx	Wx-6672-01	1,2,4	Wx
BRS Amansa Burro	3	NWx	SP-1378-66	1	NWx	Wx-6683-01	1	Wx
BRS Caipira	1,3	NWx	SP-1662-13	1	NWx	Wx-6683-03	1	Wx
BRS CS01	4	NWx	SP-1662-32	1	NWx	Wx-6703-01	4	Wx
BRS Dourada	1,3	NWx	SP-1662-41	1	NWx	Wx-6705-01	5	Wx
BRS Eucalipto	1,2,3	NWx	SP-1662-42	1	NWx	Wx-6895-05	1	Wx
BRS Formosa	1,2,4,3	NWx	SP-1662-43	1	NWx	Wx-6896-04	1,2,4	Wx
BRS Gema de Ovo	1,3	NWx	SP-1662-48	1	NWx	Wx-6896-05	1,2,4	Wx
BRS Kiriris	1,2,3	NWx	SP-1662-49	1	NWx	Wx-6958-01	5	Wx
BRS Mulatinha	1,2,4	NWx	SP-1662-65	1	NWx	Wx-6965-01	4	Wx
BRS Poti Branca	1,2,4	NWx	SP-1819-13	1	NWx	Wx-6979-01	1,2	Wx
BRS Tapioqueira	1,2,3	NWx	SP-IAC-90-06	1	NWx	Wx-7020-01	4	Wx
BRS Verdinha	1,3	NWx	SP-IAC-90-08	1	NWx	Wx-7020-02	5	Wx
Cacau	3	NWx	Tailândia	1	NWx	Wx-7025-01	1,2,4	Wx
Cascuda	4	NWx	Vassoura Preta	1,2,3	NWx	Wx-7027-01	1,2,4	Wx
Cigana Preta	1,2	NWx	Wx-6357-01	1	Wx	Wx-7032-01	5	Wx
Corrente	1,3	NWx	Wx-6359-06	1	Wx	Wx-7056-01	1,2,4	Wx
Corretão	1	NWx	Wx-6383-01	4	Wx	Wx-7099-01	1,2,4	Wx
Fécula Branca	4	NWx	Wx-6413-01	5	Wx	Wx-7153-01	4	Wx
IAC-90	1,2,3,4	NWx	Wx-6421-02	1	Wx	Wx-7424-01	2	Wx
Nega Maluca	1	NWx	Wx-6442-01	4	Wx	Wx-7424-03	1,2,4	Wx
Olho Junto	4	NWx	Wx-6454-02	1,2	Wx	Wx-7424-04	4	Wx
Salangor	1,3	NWx	Wx-6460-02	4	Wx	Wx-7429-01	1,2	Wx
SP-0061-02	1	NWx	Wx-6466-03	4	Wx	Wx-7442-04	4	Wx
SP-0061-03	1	NWx	Wx-6477-02	4	Wx	Wx-7467-01	1,2	Wx
SP-0162-02	1	NWx	Wx-6483-01	1,2	Wx	Wx-7474-01	4	Wx
SP-0222-11	1	NWx	Wx-6494-01	5	Wx	Wx-7565-01	4	Wx
SP-0222-22	1	NWx	Wx-6494-02	4	Wx	Wx-7676-01	1	Wx
SP-0222-28	1	NWx	Wx-6502-01	1,2,4	Wx	Wx-7698-16	5	Wx
SP-0222-61	1	NWx	Wx-6502-08	1,2,4	Wx	Wx-7699-06	5	Wx
SP-0222-71	1	NWx	Wx-6502-09	1,2,4	Wx	Wx-7707-14	1	Wx
SP-0222-82	1	NWx	Wx-6506-01	5	Wx	Wx-7732-07	1	Wx
SP-0222-91	1	NWx	Wx-6507-01	4	Wx	Wx-7732-27	1	Wx
SP-0728-24	1	NWx	Wx-6507-02	4	Wx	Wx-7733-16	1	Wx
SP-0728-30	1	NWx	Wx-6515-01	5	Wx	Wx-7734-02	4	Wx

Tabela S1: Cont....

Acesso	Local	Tipo	Acesso	Local	Tipo	Acesso	Local	Tipo
Wx-7736-07	4	Wx	Wx-7812-03	4	Wx	Wx-7992-05	1,2,4	Wx
Wx-7738-04	4	Wx	Wx-7813-01	1,2,4	Wx	Wx-7993-02	5	Wx
Wx-7739-04	4	Wx	Wx-7814-02	4	Wx	Wx-7999-03	4	Wx
Wx-7739-07	4	Wx	Wx-7821-01	4	Wx	Wx-8014-07	4	Wx
Wx-7742-01	4	Wx	Wx-7829-03	1,2,4	Wx	Wx-8028-03	5	Wx
Wx-7742-03	5	Wx	Wx-7843-01	5	Wx	Wx-8029-01	1,2,4	Wx
Wx-7745-05	4	Wx	Wx-7849-01	4	Wx	Wx-8029-02	4	Wx
Wx-7747-01	4	Wx	Wx-7850-01	5	Wx	Wx-8034-01	1	Wx
Wx-7748-16	4	Wx	Wx-7853-01	1,2,4	Wx	Wx-8034-02	1,2,4	Wx
Wx-7750-01	4	Wx	Wx-7859-02	1	Wx	Wx-8037-02	4	Wx
Wx-7751-01	4	Wx	Wx-7860-01	4	Wx	Wx-8039-01	1,2,4	Wx
Wx-7752-02	5	Wx	Wx-7864-07	4	Wx	Wx-8044-02	4	Wx
Wx-7754-03	4	Wx	Wx-7867-04	4	Wx	Wx-8046-02	5	Wx
Wx-7754-04	4	Wx	Wx-7868-09	4	Wx	Wx-8057-01	4	Wx
Wx-7757-02	4	Wx	Wx-7874-02	4	Wx	Wx-8058-02	5	Wx
Wx-7760-01	1,2	Wx	Wx-7877-01	4	Wx	Wx-8066-04	1,2	Wx
Wx-7769-01	1,2,4	Wx	Wx-7878-18	4	Wx	Wx-8093-03	4	Wx
Wx-7771-07	5	Wx	Wx-7882-02	4	Wx	Wx-8093-04	4	Wx
Wx-7773-04	4	Wx	Wx-7899-01	4	Wx	Wx-8104-01	1,2	Wx
Wx-7773-06	4	Wx	Wx-7905-02	1	Wx	Wx-8109-03	1,2,4	Wx
Wx-7774-03	4	Wx	Wx-7906-01	4	Wx	Wx-8109-04	1,2,4	Wx
Wx-7775-01	1,2	Wx	Wx-7909-05	1,2	Wx	Wx-8109-05	1,2,4	Wx
Wx-7775-04	4	Wx	Wx-7909-06	4	Wx			
Wx-7776-05	1	Wx	Wx-7915-01	4	Wx			
Wx-7779-06	4	Wx	Wx-7921-01	4	Wx			
Wx-7784-01	5	Wx	Wx-7924-01	5	Wx			
Wx-7784-06	4	Wx	Wx-7929-02	1	Wx			
Wx-7784-11	5	Wx	Wx-7940-01	5	Wx			
Wx-7784-13	4	Wx	Wx-7942-03	1,2	Wx			
Wx-7788-04	4	Wx	Wx-7945-01	1	Wx			
Wx-7788-07	4	Wx	Wx-7945-04	1	Wx			
Wx-7788-08	1,2	Wx	Wx-7949-01	5	Wx			
Wx-7791-07	4	Wx	Wx-7949-06	5	Wx			
Wx-7791-08	4	Wx	Wx-7950-03	4	Wx			
Wx-7791-11	1,2,4	Wx	Wx-7951-02	5	Wx			
Wx-7795-01	5	Wx	Wx-7953-02	4	Wx			
Wx-7797-04	1,2,4	Wx	Wx-7963-03	1,2	Wx			
Wx-7799-02	4	Wx	Wx-7971-01	1,2	Wx			
Wx-7802-03	5,4	Wx	Wx-7982-01	5	Wx			
Wx-7811-06	1,2,4	Wx	Wx-7992-03	4	Wx			

Local: 1 – Rio de Areia – Laje (Bahia); 2 – Novo Horizonte – Laje (Bahia); 3 – Cruz das Almas (Bahia); 4 – Dourados (Mato Grosso do Sul); 5 – Londrina (Paraná).

CONSIDERAÇÕES FINAIS

Estudos sobre o controle genético do fenótipo *waxy* e sua exploração para o desenvolvimento de variedades cerosas são relativamente recentes na cultura da mandioca. Desenvolver metodologias que possibilitem a seleção indireta e precoce são, portanto, necessárias com vistas a redução no ciclo de melhoramento para esta característica. Desta forma, nossos resultados representam avanços importantes na identificação precoce de genótipos com amido *waxy* na cultura.

No primeiro artigo, apenas a deleção de uma citosina no éxon 6 do gene *Ganule-bound starch synthase* - GBSSI (MeWxE6-del-C) discriminou o fenótipo *waxy*, enquanto que demais SNPs previamente associados ao fenótipo *waxy* foram considerados apenas variantes causais, sem estreita associação com a característica. A avaliação do gene GBSSI em uma população maior e mais diversa demonstrou que os SNPs identificados não discriminam completamente o amido *waxy*. Foi possível também determinar que, para a mutação MeWxE6-del-C, não existem genótipos heterozigóticos no germoplasma avaliado e, portanto, este marcador não deve ser utilizado para seleção assistida em diferentes *backgrounds* genéticos.

No segundo artigo, foi utilizado um painel de diversidade ainda maior para identificar variantes moleculares que possam estar associadas ao fenótipo *waxy*, com base em SNPs distribuídos em todo o genoma da mandioca. Mais de 20k de marcadores foram utilizados via GWAS na detecção de variantes associadas ao amido tipo *waxy*. Dez variantes foram associadas e co-localizadas em regiões gênicas. Cinco destas estão localizados em genes relacionados a expressão gênica e estudos adicionais serão necessários para avaliar seu efeito sobre o fenótipo. Também discutimos brevemente a relação dos genes co-localizados com a rota metabólica do amido. Os resultados ainda indicam a possibilidade de interações epistáticas ou aditivas entre as variantes associadas, que devem ser melhor investigadas.

No terceiro artigo, foi explorado o potencial de classificação do genótipo *waxy*, por avaliação indireta da folha por espectrometria do infravermelho próximo com transformada de Fourier (FT-NIR). Diversos modelos supervisionados para classificação e predição do fenótipo em mais de 296 genótipos coletados em diferentes localidades, demonstraram que *Bayesian generalized linear model* - BGLM e *support vector machines with linear kernel* – SVM foram mais eficazes na

classificação correta dos fenótipos (93%) na validação externa, sendo, portanto, promissores para a seleção precoce de genótipos portadores do amido *waxy*.