

**HELIO DE SOUZA JUNIOR**

**DESENVOLVIMENTO E VALIDAÇÃO DE SNPs PARA SELEÇÃO ASSISTIDA  
POR MARCADORES EM BATATA-DOCE HEXAPLOIDE**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

Orientador: Guilherme da Silva Pereira

**VIÇOSA - MINAS GERAIS  
2024**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

S729d  
2024 Souza Junior, Helio de, 1992-  
Desenvolvimento e validação de SNPs para seleção  
assistida por marcadores em batata-doce hexaploide / Helio de  
Souza Junior. – Viçosa, MG, 2024.  
1 dissertação eletrônica (50 f.): il. (algumas color.).

Orientador: Guilherme da Silva Pereira.  
Dissertação (mestrado) - Universidade Federal de Viçosa,  
Departamento de Agronomia, 2024.  
Referências bibliográficas: f. 45-50.  
DOI: <https://doi.org/10.47328/ufvbbt.2024.623>  
Modo de acesso: World Wide Web.

1. Batata-doce - Melhoramento genético. 2. Marcadores  
genéticos. I. Pereira, Guilherme da Silva, 1986-. II. Universidade  
Federal de Viçosa. Departamento de Agronomia. Programa de  
Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 635.222

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317


**HELIO DE SOUZA JUNIOR**

**DESENVOLVIMENTO E VALIDAÇÃO DE SNPs PARA SELEÇÃO ASSISTIDA  
POR MARCADORES EM BATATA-DOCE HEXAPLOIDE**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.


APROVADA: 7 de junho de 2024.

Assentimento:

Documento assinado digitalmente  
 **HELIO DE SOUZA JUNIOR**  
Data: 01/10/2024 06:36:24-0300  
Verifique em <https://validar.iti.gov.br>

---

**Helio de Souza Junior**  
Autor

Documento assinado digitalmente  
 **GUILHERME DA SILVA PEREIRA**  
Data: 01/10/2024 08:23:04-0300  
Verifique em <https://validar.iti.gov.br>

---

**Guilherme da Silva Pereira**  
Orientador

*A Deus, meu companheiro SEMPRE.  
Aos meus pais, minha admiração e meus exemplos para toda a vida. Ao meu querido irmão  
Leandro de Souza (in memoriam) e para minha sobrinha amada Ana Beatriz (in memoriam).  
Tudo por vocês, tudo para vocês.*

## AGRADECIMENTOS

A Deus pela vida, saúde e pelas pessoas maravilhosas que colocou no meu caminho.

À Universidade Federal de Viçosa (UFV) e ao Curso de Pós-Graduação em Genética e Melhoramento pelo ensino de excelência que me proporcionaram.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos durante o curso.

Ao *Centro Internacional de la Papa (CIP)* e ao *National Crops Resources Research Institute (NaCRR)*, seus pesquisadores e técnicos, por gerarem e disponibilizarem os dados genotípicos e fenotípicos utilizados neste trabalho.

Ao Prof. Dr. Guilherme da Silva Pereira, pela orientação, pelos conhecimentos transmitidos para o meu crescimento e desenvolvimento, por todas as oportunidades e, principalmente, por ser uma pessoa compreensiva nos momentos mais difíceis da minha vida.

À todos os professores que compartilharam seu valioso conhecimento através de disciplinas e cursos oferecidos pela PPGGM, em especial aos professores Aluizio Borém, Antônio Policarpo Souza Carneiro, Paulo Roberto Cecon, Leonardo Lopes Bhering e Rodrigo de Oliveira Lima pelos valiosos ensinamentos transmitidos durante o mestrado, seja nas disciplinas ou nas conversas “informais” nos corredores.

Aos pós-docs Dra. Iara Gonçalves e Dr. João Bachega, pessoas com as quais tive o prazer de conviver. Em especial a Dra. Carla Cristina da Silva, sempre disponível para conversar, aconselhar e me ajudar. Por sua boa vontade em discutir genética molecular e genética-estatística comigo. Sou realmente grato a vocês, me ensinaram muito sobre um campo da ciência que eu pouco conhecia.

A todos os colegas que passaram pelo Laboratório de Bioinformática e do Laboratório de Biotecnologia e Melhoramento Vegetal, que não vou arriscar-me a nomeá-los para que não cometa nenhuma injustiça, por proporcionarem momentos de descontração e aprendizado.

Aos meus grandes amigos Carlos Henrique, Maria Rocha, Prof. Deonísio Destro, Dr. Assis, Profa. Luzia Paccola, Prof. Ricardo Tadeu, Tainá Miranda, Potira Miranda e João Franco pela amizade e pela torcida constante.

A todos os meus colegas que eu fiz no curso, em especial a Lisandra Moura e Fabrício Rainha Ribeiro (*in memoriam*), pela amizade sincera, momentos de descontração e pelo companheirismo apresentado em todo o mestrado.

Ao meu primo Cristiano Rodrigues (*in memoriam*), a Profa. Jurdilei Paccola (*in memoriam*) e ao meu amigo Renan Cantanti Marques (*in memoriam*) - você era um grande microbiologista -, obrigado por todo carinho e pelas longas conversas.

À minha querida namorada Joyce Araújo, obrigado pela força, pelo carinho e principalmente pela paciência.

Por fim, mas não menos importante, aos meus pais Helio de Souza e Maria Sueli Teixeira da Silva, com todo meu amor e gratidão, por tudo que fizeram por mim ao longo de minha vida. Aos meus irmãos e minhas irmãs, em especial para o meu querido irmão Leandro de Souza (*in memoriam*) que sempre acreditou em mim e para minha sobrinha amada Ana Beatriz (*in memoriam*). À toda minha família, pela torcida constante e amor incondicional mesmo à 1.350 km distante de casa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) – Código de Financiamento APQ-02529-22.

**De coração, meu muito obrigado!**

*“Não importa o que a vida fez de você. Importa o que  
você fez com que a vida fez de você.”*

(Jean-Paul Sartre)

## RESUMO

SOUZA JUNIOR, Helio de. M.Sc., Universidade Federal de Viçosa, junho de 2024. **Desenvolvimento e validação de SNPs para seleção assistida por marcadores em batata-doce hexaploide.** Orientador: Guilherme da Silva Pereira.

A batata-doce é uma importante cultura do mundo, sendo considerada de segurança alimentar e nutricional na África Subsaariana. Trata-se de uma espécie autohexaploide, de alogamia obrigatória e com um elevado grau de heterozigiosidade, fatores que tornam desafiadoras análises moleculares envolvendo locos de herança quantitativa (QTLs). O melhoramento de cultivares de batata-doce com características desejáveis pode ser facilitado e acelerado utilizando-se da seleção assistida por marcadores, ferramenta da qual outras espécies, sobretudo diploides, já se beneficia. O objetivo deste trabalho foi desenvolver marcadores baseados em polimorfismos de único nucleotídeo (SNPs) presentes em QTLs previamente identificados em populações de mapeamento da espécie. Um total de 60 SNPs foram selecionados nos cromossomos 3, 12 e 13 em genes putativos presentes em regiões genômicas associadas aos seguintes caracteres: conteúdo de matéria seca, teores de amido, glicose, sacarose,  $\beta$ -amilase e  $\beta$ -caroteno, e cores da polpa da casca. A metodologia de genotipagem *kompetitive allele specific PCR* (KASP) foi utilizada para obtenção das intensidades dos sinais dos alelos, *A* e *B*, dos SNPs em 86 amostras de dialelo 8×8, com fenótipos disponíveis, mais sete cultivares. A dosagem alélica foi inferida a partir de modelos de mistura ajustados utilizando o pacote R fitPoly v.3.0.0. Modelos de regressão linear simples entre os valores fenotípicos e as dosagens alélicas foram ajustados para 51 SNPs polimórficos codificados para aditividade e dominância ou baseados na razão  $B/(A + B)$ . De acordo com a estatística *F*, os coeficientes de regressão foram significativos ( $P < 0,001$ ), para cinco, onze, sete e dois SNPs para conteúdo de matéria seca, cor da polpa, cor da casca e firmeza de raiz (correlacionado com o conteúdo de  $\beta$ -amilase), respectivamente. A análise de regressão demonstrou que para firmeza de raiz, houve maior significância para modelo de dominância de snpIB00118 ( $R^2 = 0,20$ ). Para os demais caracteres, modelos de aditividade ou baseados em razão apresentaram maior significância, destacando-se snpIB00070 ( $R^2 = 0,22$  ou  $R^2 = 0,20$ , respectivamente) para conteúdo matéria seca, snpIB00083 ( $R^2 = 0,28$  ou  $R^2 = 0,29$ ) para cor de polpa, e snpIB00094 ( $R^2 = 0,46$  ou  $R^2 = 0,49$ ) para cor de casca. O uso de razões de intensidades alélicas demonstrou similar poder para detectar associação quando comparado aos modelos aditivos, facilitando a análise. No entanto, essa abordagem não capturou associações no caso de controle genético dominante, como, possivelmente, no caso da firmeza. Nas análises de regressão linear múltipla, seis, cinco, seis e dois SNPs foram selecionados para conteúdo de matéria seca, cor de polpa, cor de casca e firmeza, com acréscimos significativos na proporção da variação explicada de todos os caracteres exceto firmeza. Os SNPs com maiores associações explicam parcial e significativamente a variação fenotípica de caracteres de interesse, e podem ser utilizadas para a aplicação imediata da seleção assistida para o melhoramento genético desses caracteres.

Palavras-chave: *Ipomoea batatas*. KASP. Seleção assistida por marcadores. Regressão linear. QTL.

## ABSTRACT

SOUZA JUNIOR, Helio de. M.Sc., Universidade Federal de Viçosa, June, 2024. **Development and validation of SNPs for marker-assisted selection in hexaploid sweetpotato.** Adviser: Guilherme da Silva Pereira.

Sweetpotato is an important crop worldwide, and it is considered a food security crop in sub-Saharan Africa. As an autohexaploid, outbred, highly heterozygous crop, molecular analysis and quantitative trait loci (QTL) identification are challenging. The improvement of sweetpotato varieties with desirable traits can be boosted through marker-assisted selection, a tool that other species, mostly diploids, have already benefited from. The aim of this work was to develop single nucleotide polymorphism (SNP) markers derived from QTL regions previously identified in sweetpotato mapping populations. A total of 60 SNPs were selected on chromosomes 3, 12 and 13 in genes found to be putatively associated to the following traits: dry matter, starch,  $\beta$ -carotene, glucose, sucrose,  $\beta$ -amylase contents, and flesh and skin colors. Kompetitive allele specific PCR (KASP) genotyping methodology was used for measuring alleles *A* and *B* signals from 93 samples from an 8×8 diallel. Allele dosage was inferred using mixture models as implemented in R package fitPoly v.3.0.0. Simple linear regression models between the phenotypic values and the dosages of allelic intensities were fitted for 51 polymorphic SNPs for each additive, dominance, and ratio  $B/(A + B)$  based models. According to the *F*-statistics, regression coefficients were significant ( $P < 0.001$ ) for five, eleven, seven, and two SNPs for dry matter content, flesh color, skin color, and firmness (negatively correlated to  $\beta$ -amylase content), respectively. Regression analysis for firmness showed that there was a significant dominance model for snpIB00118 ( $R^2 = 0.20$ ). For the other traits, both additive and ratio-based models were significant for snpIB00070 ( $R^2 = 0.22$  or  $R^2 = 0.20$ , respectively) for dry matter content, snpIB00083 ( $R^2 = 0.28$  or  $R^2 = 0.29$ ) for flesh color, snpIB00094 ( $R^2 = 0.46$  or  $R^2 = 0.49$ ) for skin color. The use of allelic intensity ratios facilitates the analysis since allele dosage classification is not needed. However, this approach did not show associations in the case of dominant genetic control, as it is possibly the case of firmness. Multiple linear regression models with six, five, six, and two SNPs selected for dry matter content, flesh color, skin color, and firmness, respectively, have shown significant increase in proportion of explained variation for all traits except firmness. The SNPs with the highest associations explain partially and significantly the phenotypic variation of the traits of interest and can be used for marker-assisted selection purposes.

Keywords: *Ipomoea batatas*. KASP. Marker-assisted selection. Linear regression. QTL.

## LISTA DE FIGURAS

- Figura 1 – Esquema geral de um típico programa de melhoramento convencional de batata-doce. Círculos representam plantas individuais enquanto quadrados representam múltiplas plantas do mesmo clone. O número de anos e clones por fase depende do programa de melhoramento em questão ..... 13
- Figura 2 – Gráfico de dispersão das intensidades de sinal dos alelos  $A$  e  $B$  para 60 KASPs. As cores exibem a classificação fornecida pela empresa que originou os dados. ‘Branco’ significa que não foi adicionada amostra de DNA à reação, enquanto ‘baixo sinal’ e ‘não classificado’ quer dizer que os pontos não receberam qualquer uma das três possíveis classes pseudo-diploides,  $AA = 0$ ,  $AB = 1$  ou  $BB = 2$ ..... 30
- Figura 3 – Classificação obtida a partir da função *saveMarkerModels* do pacote do R *fitPoly*. As cores referem-se às doses mais prováveis enquanto a transparência refere-se ao valor de  $P$  associadas às doses. Os dados foram filtrados para eliminar os ‘brancos’ e as amostras contendo baixo sinal. Dados perdidos (‘NA’) resultaram de falha na chamada ..... 31
- Figura 4 – Classificação obtida a partir da função *CodomMarker* do pacote do R *fitPoly* mais ajustes manuais. As cores referem-se às doses mais prováveis enquanto a transparência refere-se ao valor de  $P$  associadas às doses. Os dados foram filtrados para eliminar os ‘brancos’ e as amostras contendo baixo sinal. Dados perdidos (‘NA’) resultaram de falha na chamada ..... 32
- Figura 5 – Comparação de resultados da dosagem alélica usando funções do *fitPoly* *saveMarkerModels* (A), *CodomMarker* (B) e correção manual (C). Os marcadores *snpIB00069* e *snpIB00072* apresentaram melhores resultados a partir das funções *saveMarkerModels* e *CodomMarker*, respectivamente. Os marcadores *snpIB00070*, *snpIB00074* e *snpIB00094* exigiram correção manual ..... 33
- Figura 6 – Classificação final em classes de doses alélicas obtidas a partir do *fitPoly*. As cores referem-se às doses mais prováveis enquanto a transparência refere-se ao valor de  $P$  associadas às doses. Os dados foram filtrados para eliminar os ‘brancos’ e as amostras contendo baixo sinal. Dados perdidos (‘NA’) resultaram de falha na chamada ..... 34
- Figura 7 – Valores de  $P$ , transformados usando  $-\log_{10}$ , para os 51 KASPs testados contra quatro caracteres em população diversa de batata-doce ( $N = 86$ ) considerando quatro modelos de codificação dos genótipos. Os coeficientes de determinação  $R^2$  a partir de modelos de regressão simples são expressos em gradiente de cor. Efeitos considerados significativos ( $P < 0,001$ ) ficam acima da linha tracejada. Os KASPs estão agrupados de acordo com os genes anotados no genoma de *Ipomoea trifida* v.3, a partir dos quais SNPs-alvo foram selecionados ..... 38
- Figura 8 – Desequilíbrio de ligação a partir das correlações ao quadrado ( $r^2$ ) entre as dosagens alélicas de 51 marcadores KASP genotipados em clones de batata-doce ( $N = 93$ ) ..... 39
- Figura 9 – Regressão linear considerando quatro modelos de codificação dos genótipos para os dois SNPs com efeitos mais significativos para cada caráter: matéria seca (A), cor da polpa (B), cor da casca (C), e firmeza (D). Nos modelos ‘Dom. REF’ e ‘Dom. ALT’, os genótipos homocigóticos para os respectivos alelos  $A$  e  $B$  (classe 0) foram testados contra o conjunto dos demais genótipos (classe 1). Os números de indivíduos são mostrados sob cada classe ..... 40

## LISTA DE TABELAS

Tabela 1 – Estatísticas resumidas dos QTLs para caracteres de qualidade alvos do presente estudo em populações de mapeamento envolvendo ‘Beauregard’ (B) e ‘Tanzania’ (T). $h_{QTL}^2$ é a herdabilidade do QTL, e o SNP mais próximo refere-se à posição no genoma de <i>Ipomoea trifida</i> v.3 .....	17
Tabela 2 – Descrição dos genitores da população <i>Mwanga Diversity Panel</i> provenientes de duas populações denominadas A e B quanto ao país de origem e cor da polpa .....	18
Tabela 3 – Pedigree dos 93 genótipos de batata-doce genotipados para validação dos marcadores KASPs. ‘NAROSPOT 1’ foi amostrado em duplicata, e ‘NA’ indica informação indisponível .....	20
Tabela 4 – Informação sobre SNPs convertidos em KASPs. Anotação e posição referem-se ao genoma de <i>Ipomoea trifida</i> v.3.....	26
Tabela 5 – Marcadores KASPs com efeitos significativos ( $P < 0,001$ ) para os caracteres conteúdo de matéria seca, cor de polpa, cor de casca e firmeza para diferentes modelos de codificação da informação do SNP, e seus respectivos interceptos e coeficiente de determinação ( $R^2$ ).....	37
Tabela 6 – Estimativas dos interceptos, efeitos e valores de P associados com marcadores KASP combinados em modelos de regressão linear múltipla para cada caráter, e seus respectivos coeficientes de determinação múltiplo ( $R^2$ ) e múltiplo ajustado ( $R_{aj}^2$ ). Os SNPs foram selecionados a partir de procedimento <i>stepwise</i> utilizando critério de informação Bayesiano e a significância foi expressa como *** $P < 0,001$ , ** $P < 0,01$ e * $P < 0,05$ .....	43

## SUMÁRIO

1. INTRODUÇÃO.....	12
2. MATERIAIS E MÉTODOS.....	16
2.1. Seleção dos SNPs.....	16
2.2. Material vegetal .....	18
2.3. Dados genotípicos.....	19
2.4. Dados fenotípicos .....	21
2.4.1. Conteúdo de matéria seca, cor da polpa e cor da casca .....	21
2.4.2. Firmeza da raiz.....	22
2.5. Regressão linear simples e múltipla.....	22
3. RESULTADOS E DISCUSSÃO.....	25
3.1. Desenvolvimento dos SNPs.....	25
3.2. Dosagem alélica .....	27
3.2. Validação dos marcadores .....	35
4. CONCLUSÕES .....	44
REFERÊNCIAS .....	45

## 1. INTRODUÇÃO

A batata-doce, *Ipomoea batatas* (L.) Lam., é a sétima cultura alimentícia em importância no mundo devido ao seu valor nutricional e como fonte de renda para pequenos agricultores. Em áreas com forte insegurança alimentar, como a África Subsaariana, a América Latina e países asiáticos, a batata-doce se torna alimento-chave para a subsistência de populações nestas regiões (MOTSA; MODI; MABHAUDHI, 2015; MUKHOPADHYAY *et al.*, 2011). As raízes de reserva são ricas em amido, vitaminas C, E e do complexo B, fibras, ferro, zinco, potássio e magnésio. Além disso, as raízes de polpa laranja são ricas em  $\beta$ -caroteno, precursores da vitamina A, enquanto as de polpa roxa possuem altos teores de antocianinas, poderosos antioxidantes (BOVELL-BENJAMIN, 2007; WOOLFE, 1992). Na África Subsaariana, a introdução de variedades biofortificadas de polpa laranja está ajudando no combate à fome, má nutrição e deficiência de vitamina A (LOW *et al.*, 2017).

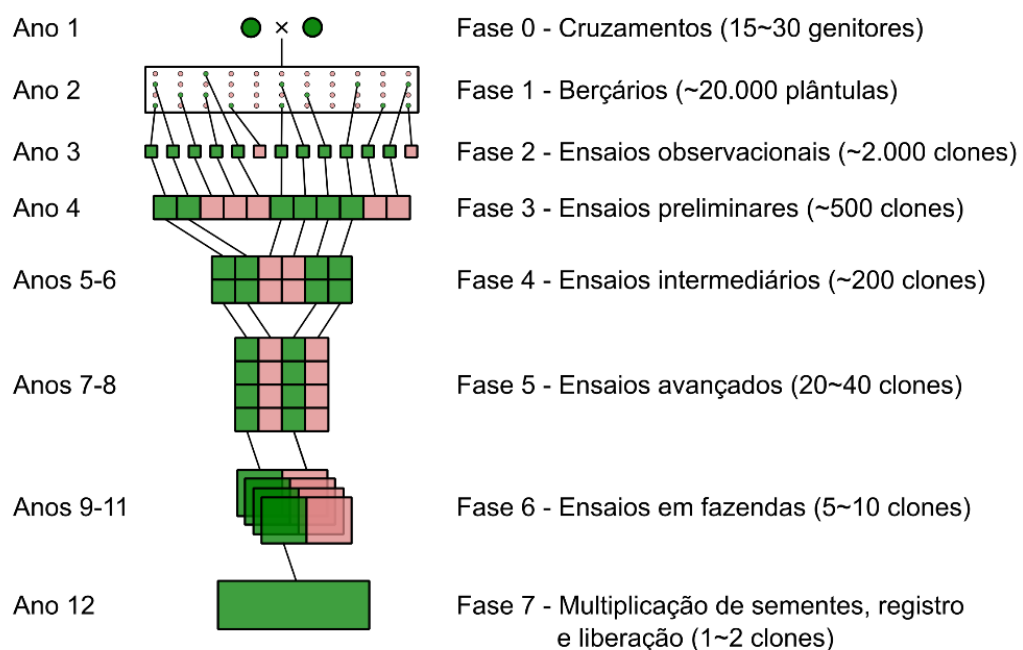
*Ipomoea batatas* tem origem nos trópicos e é plantada praticamente em todas as regiões tropicais e subtropicais. Em 2022, a produção mundial total da cultura foi de aproximadamente 86,5 milhões de toneladas. A China é o maior país produtor, respondendo por 53,6% da produção mundial, enquanto a África responde por 34,2%. As Américas são responsáveis por apenas 3,7% da produção mundial de batata-doce, com os Estados Unidos (9º maior produtor) e o Brasil (14º maior produtor) produzindo 1,2 milhão e 847,1 mil toneladas (t), respectivamente (FAO, 2023). A região Nordeste é a maior produtora brasileira (351.623 t), seguida pelas regiões Sudeste (242.777 t) e Sul (232.737 t). Entre os estados brasileiros, São Paulo é o maior produtor de batata-doce com 160.514 t produzidos na safra 2022, seguido pelo Rio Grande do Sul (155.070 t) e Ceará (116.668 t) (IBGE, 2023).

Ainda que possa responder bem a tratamentos agrícolas, batata-doce é tipicamente relatada como uma cultura de fácil manejo e baixo custo de produção, adaptada ao clima tropical e subtropical. Plantas dessa espécie possuem uma alta variabilidade genética que favorece as possibilidades de aprimoramento da cultura, para caracteres bióticos e abióticos, incluindo conteúdo de matéria seca e produtividade. Existem diversas cultivares dessa espécie, as quais podem ser classificadas levando em consideração, dentre outras, as seguintes características: cor interna e externa, tamanho e formato das raízes de armazenamento, cor de folhas e flores (VARGAS *et al.*, 2022). Caracteres de qualidade, como cor da casca, cor da polpa e textura, apresentam ampla variabilidade natural e se relacionam com preferências de variedades pelo consumidor (KATAYAMA *et al.*, 2017; NAKITTO *et al.*, 2024).

A batata-doce é uma espécie autohexaploide ( $2n = 6x = 90$ ), com genoma relativamente grande (aproximadamente 3 Gb), e de alogamia obrigatória, com elevado grau de heteroziguidade. Combinadas, essas características conferem à espécie alta complexidade genômica e dificultam a aplicação de ferramentas genético-genômicas no melhoramento genético da cultura (SRISUWAN; SIHACHAKR; SILJAK-YAKOVLEV, 2006; WU *et al.*, 2018). A fecundação cruzada obrigatória dá-se devido à ocorrência de incompatibilidade, de tal modo que a obtenção de linhagens endogâmicas é extremamente difícil. O melhoramento genético da cultura é todo baseado em progênies  $F_1$ , portanto diferentes entre si, resultantes de cruzamentos de genitores heterozigóticos compatíveis, cada qual com potencial de se tornar uma cultivar (GRÜNEBERG *et al.*, 2022). Os indivíduos são clonados e testados sob diferentes densidades, com aumento progressivo de número de plantas por parcela, repetições dentro de ambiente e de ambientes avaliados à medida que as fases avançam (Figura 1). Um ciclo de seleção pode levar mais de oito anos para se concretizar (MWANGA *et al.*, 2021), a depender de quantas avaliações no tempo são realizadas por determinado programa. Desta forma, o desenvolvimento de técnicas de otimização da seleção de clones promissores para a cultura, que possibilitem encurtar o ciclo de melhoramento, se faz premente.

Marcadores moleculares, desenvolvidos a partir de variações na sequência de DNA entre indivíduos, são uma ferramenta genética com inúmeras aplicações em programas de

Figura 1 – Esquema geral de um típico programa de melhoramento convencional de batata-doce. Círculos representam plantas individuais enquanto quadrados representam múltiplas plantas do mesmo clone. O número de anos e clones por fase depende do programa de melhoramento em questão



Fonte: Adaptado de Andrade *et al.* (2017).

melhoramento de diversas culturas. Eles são usados para caracterização de germoplasma, avaliação de diversidade genética, acompanhamento de introgressão de genes, confirmação de parentesco, *fingerprinting* e proteção de cultivares, e seleção assistida por marcadores (SAM) (BOHAR; CHITKINENI; VARSHNEY, 2020). Uma das aplicações mais importantes dos marcadores moleculares é o mapeamento de locos de características quantitativas (do inglês, *quantitative trait loci* – QTLs). A maioria dos caracteres de interesse agrônômico apresentam controle poligênico, sendo que o mapeamento de QTLs pode contribuir com o entendimento do controle genético de tais características, com a identificação de genes candidatos para o melhoramento genético e de marcadores com potencial para SAM. Em poliploides, entretanto, o mapeamento de QTLs é dificultado pela homologia entre regiões cromossômicas, o tipo de segregação de genes/marcadores e um alto número de genótipos por loco. Devido a estas dificuldades, a maioria dos marcadores são classificados como dominantes (alelos de dose única), para os quais a informação de dosagem alélica (número de cópias de cada alelo) dos locos não pode ser utilizada. Felizmente, nos últimos anos, o desenvolvimento de métodos moleculares e genético-estatísticos específicos para lidar com a complexidade de espécies autoploides tem avançado (MOLLINARI; SERANG, 2015; PEREIRA *et al.*, 2023).

Diferentes tipos de marcadores moleculares são utilizados para o mapeamento de QTLs para resistência a doenças, produtividade e caracteres qualitativos em batata-doce, sendo os polimorfismos de base única (do inglês, *single nucleotide polymorphisms* – SNPs) os mais utilizados recentemente (GEMENET *et al.*, 2020; OLOKA *et al.*, 2021). Os SNPs são variações em um único nucleotídeo em um ponto específico do DNA, sendo os polimorfismos de DNA mais abundantes nos genomas. Eles são marcadores bialélicos altamente informativos de natureza codominante, isto é, ambos os alelos podem ser identificados. Desta maneira, estudos em espécies poliploides, como a batata-doce, se beneficiam com uso deste marcador em virtude da possibilidade de caracterizar a dosagem alélica (PEREIRA *et al.*, 2023). Entretanto, apesar dos avanços no desenvolvimento de marcadores moleculares e na identificação de regiões genômicas que influenciam caracteres de importância para a cultura, o uso dos marcadores na rotina dos programas de melhoramento de batata-doce ainda não foi implementado.

A descoberta de SNPs tipicamente acontece após sequenciamento e alinhamento de sequências de DNA. Uma vez conhecida a variação do tipo SNP, há diversas maneiras de acessá-la em um conjunto de indivíduos. A reação em cadeia da polimerase (do inglês, *polymerase chain reaction* – PCR) quantitativa tem sido empregada para a detecção destes SNPs utilizando, principalmente, sondas de hidrólise pela metodologia *TaqMan*. No entanto,

metodologias que empregam apenas o uso de *primers* já foram desenvolvidas, como a metodologia denominada *kompetitive allele specific PCR* (KASP) (SEMAGN *et al.*, 2014).

A técnica KASP se baseia na transferência de energia por ressonância de fluorescência (FRET) para geração de sinal (SEMAGN *et al.*, 2014). No entanto, o método emprega dois *primers forward* (diretos) alelo-específicos, um *primer reverse* (reverso) comum, e um par de oligonucleotídeos marcados complementares às caudas dos *primers forward*. Os oligos marcados emitem sinal apenas quando incorporados aos amplicons correspondentes aos *primers forward* alelo-específicos, gerando um sinal mensurável (MEADE *et al.*, 2020). Se o genótipo em um determinado SNP for homocigoto, apenas um dos dois possíveis sinais fluorescentes será gerado. Se o genótipo for heterocigoto, um sinal fluorescente misto será produzido. Em indivíduos poliploides, a dosagem pode ser obtida a partir da intensidade relativa de cada sinal.

Dosagem em indivíduos poliploides constitui uma extensão do mesmo conceito em diploides. Em espécies diploides, como a mandioca (*Manihot esculenta* Crantz), pode haver até três classes genóticas correspondendo à contagem do alelo de alternativo *B*, isto é,  $AA = 0$ ,  $AB = 1$ ,  $BB = 2$ . Em espécies tetraploides, como a batata (*Solanum tuberosum* L.), até cinco classes genóticas podem ser observadas:  $AAAA = 0$ ,  $AAAB = 1$ ,  $AABB = 2$ ,  $ABBB = 3$ ,  $BBBB = 4$ . Em espécies hexaploides, como a batata-doce, podem existir até sete classes genóticas:  $AAAAAA = 0$ ,  $AAAAAB = 1$ ,  $AAAABB = 2$ ,  $AAABBB = 3$ ,  $AABBBB = 4$ ,  $ABBBBB = 5$ ,  $BBBBBB = 6$ . Os marcadores KASP permitem pontuação bialélica em loco específico e oferecem níveis extremamente altos de robustez e precisão a um custo relativamente baixo.

Neste contexto, objetivou-se desenvolver e validar marcadores do tipo KASP a partir de SNPs localizados em regiões genômicas que controlam caracteres de qualidade em batata-doce hexaploide previamente identificados (AMANKWAAH, 2019; GEMENET *et al.*, 2020; UTOBLO *et al.*, 2018), além de estabelecer as melhores estratégias para a análise dos dados moleculares e fenotípicos. Para tanto, polimorfismos, dosagem alélica e putativa associação dos locos selecionados com caracteres de importância econômica foram avaliados em genótipos de batata-doce hexaploide de uma população de melhoramento do *International Potato Center* (CIP) estabelecida em Uganda. Os resultados gerados neste trabalho têm o potencial de serem utilizados em diferentes programas de melhoramento de batata-doce ao redor do mundo, incluindo o Brasil, para implementação de marcadores moleculares para SAM.

## 2. MATERIAIS E MÉTODOS

### 2.1. Seleção dos SNPs

Anterior ao desenvolvimento deste trabalho, duas populações biparentais foram submetidas ao mapeamento de QTLs após obtenção de mapa de ligação e avaliação fenotípica para diversos caracteres em países diferentes. Tratou-se de uma população, B×T, resultado do cruzamento entre a cultivar estadunidense ‘Beauregard’ (B), de polpa alaranjada, e a variedade local ugandense ‘Tanzania’ (T), de polpa creme, com 315 progênies, e seu cruzamento recíproco, T×B, com 248 progênies. A população B×T originou um mapa com 30.694 marcadores e 2.708 cM (MOLLINARI *et al.*, 2020), e a recíproca resultou em um mapa com 14.813 marcadores e 2.120 cM (OLOKA *et al.*, 2021).

No Peru, a população B×T foi avaliada em cinco ambientes para caracteres relacionados ao conteúdo de matéria seca, teor de amido, teor de  $\beta$ -caroteno, e cor de polpa (GEMENET *et al.*, 2020). Em Gana, uma amostra de 246~270 progênies da mesma população B×T foi caracterizada em oito ambientes para os mesmos caracteres (UTOBLO *et al.*, 2018), além de teor de glicose e cor da casca (dados não publicados). Finalmente, nos EUA, a população T×B foi avaliada em duas safras para caracteres semelhantes, além de teor de  $\beta$ -amilase (AMANKWAAH, 2019). QTLs de maiores efeitos, identificados a partir da metodologia descrita por Da Silva Pereira *et al.* (2020), com herdabilidades superiores a 0,21 (valores de  $P < 1 \times 10^{-7}$  para a estatística do teste), foram selecionados para identificação de SNPs-alvo deste trabalho (Tabela 1). Ainda que os mapas de B×T e T×B não puderam ser comparados diretamente quanto às medidas de distância genética (em centiMorgans), os SNPs que os compuseram foram mapeados usando o mesmo genoma de *Ipomoea trifida* v.3 (WU *et al.*, 2018), espécie diploide mais proximamente relacionada a *I. batatas*. Essa referência permitiu, portanto, a comparação com base nas distâncias físicas (em pares de base – pb).

Com a realização desses trabalhos (AMANKWAAH, 2019; GEMENET *et al.*, 2020; UTOBLO *et al.*, 2018), observou-se a existência de QTLs de maiores efeitos em duas regiões do cromossomo 3, duas regiões no cromossomo 12, e uma região no cromossomo 13. Na porção proximal do cromossomo 3, a 2,99~3,18 Mpb, colocalizaram-se QTLs para teor de amido e conteúdo de matéria seca, positivamente correlacionados entre si, e para teor de  $\beta$ -caroteno e cor da polpa, positivamente correlacionados entre si, e negativamente correlacionados com teor de amido e conteúdo de matéria seca (GEMENET *et al.*, 2020). Na porção distal do cromossomo 3, a 11,05~11,57 Mpb, colocalizaram-se QTLs para teor de glicose. Similarmente,

Tabela 1 – Estatísticas resumidas dos QTLs para caracteres de qualidade alvos do presente estudo em populações de mapeamento envolvendo ‘Beauregard’ (B) e ‘Tanzania’ (T).  $h_{QTL}^2$  é a herdabilidade do QTL, e o SNP mais próximo refere-se à posição no genoma de *Ipomoea trifida* v.3

Caráter	Grupo de ligação	Posição (cM)	SNP mais próximo	Valor de P	$h_{QTL}^2$
B×T (N = 315) avaliada no Peru nos anos de 2016 e 2017 (GEMENET <i>et al.</i> , 2020)					
Conteúdo de matéria seca	3	37,44	S3_3185578	<1,00e-16	0,37
Teor de amido	3	37,44	S3_3185578	<1,00e-16	0,51
Teor de β-caroteno	3	36,14	S3_2994719	<1,00e-16	0,50
	12	146,02	S12_22131994	<1,00e-16	0,29
Cor da polpa	3	36,14	S3_2994719	<1,00e-16	0,49
	12	146,02	S12_22131994	<1,00e-16	0,28
B×T (N = 246~270) avaliada em Gana nos anos de 2017 e 2018 (UTOBLO <i>et al.</i> , 2018)					
Conteúdo de matéria seca	3	37,44	S3_3185578	<2,22e-16	0,27
Teor de amido	3	37,44	S3_3185578	<2,22e-16	0,38
Teor de glicose	3	214,24	S3_11573949	<2,22e-16	0,27
Teor de β-caroteno	3	37,44	S3_3185578	<2,22e-16	0,37
	12	174,03	S12_22131994	<2,22e-16	0,25
Cor da polpa	3	36,14	S3_2994719	<2,22e-16	0,40
	12	143,25	S12_21843571	<2,22e-16	0,21
Cor da casca	12	34,23	S12_2478102	<2,22e-16	0,48
T×B (N = 248) avaliada nos EUA nos anos de 2016 e 2017 (AMANKWAAH, 2019)					
Conteúdo de matéria seca	3	29,82	S3_3104661	2,32e-09	0,28
Teor de amido	3	29,82	S3_3104661	6,44e-08	0,29
Teor de glicose	3	151,72	S3_11050542	<1,00e-16	0,28
Teor de β-caroteno	3	29,82	S3_3104661	<1,00e-16	0,45
	12	126,01	S12_22117539	<1,00e-16	0,27
Teor de β-amilase	13	79,68	S13_18484853	<1,00e-16	0,54

Fonte: Adaptado de Amankwaah (2019); Gemenet *et al.* (2020); Utoblo *et al.* (2018)

teor de β-caroteno e cor de polpa apresentaram QTLs colocalizados na porção distal do cromossomo 12, a 22,11~22,13 Mpb, enquanto QTL para cor de casca foi detectado na região proximal, a 2,48 Mpb (Utoblo *et al.* 2018, dados não publicados). Finalmente, QTL para teor de β-amilase foi detectado no cromossomo 13, a 18,48 Mbp (AMANKWAAH, 2019).

Utilizando o *genome browser* disponibilizado em <https://sweetpotato.uga.edu/> para o genoma de *I. trifida* v.3, SNPs contidos dentro de genes, nas regiões dos respectivos QTLs, putativamente relacionados aos caracteres em estudo foram listados. Os SNPs disponibilizados no *genome browser* foram detectados em trabalho anterior (WU *et al.*, 2018), que realizou o resequenciamento dos 16 genótipos de batata-doce usados nos cruzamentos que originaram a população de validação utilizada neste estudo (descrita na próxima seção). A seleção final dos SNPs relacionou-se a critérios envolvendo um conteúdo de GC entre 40~60% e um número de até 2 variantes na vizinhança ao SNP-alvo para a região flanqueadora de 50 pb de cada lado do SNP, conforme implementado no *script* disponível em <https://github.com/HPCBio/eib-marker-design>. Um total de 60 SNPs foram selecionados e enviados à empresa Intertek (Suíça) para síntese dos *primers* e realização de ensaios KASP.

## 2.2. Material vegetal

Para fins de validação dos marcadores KASP, foi utilizada uma amostra de indivíduos de uma população chamada de *Mwanga Diversity Panel* (MDP), independente das populações de mapeamento utilizadas na identificação de QTLs. A população MDP foi desenvolvida pelo Centro Internacional da Batata (CIP) e constituiu-se de um dialelo parcial obtido a partir de cruzamentos entre 8 genitores femininos e 8 genitores masculinos, resultando em 64 famílias com cerca de 30 irmãos-completos cada, totalizando 1920 indivíduos. Os 16 genótipos parentais que foram cruzados para desenvolver esta população foram provenientes de duas populações, classificadas como A (50 clones) e B (80 clones) a partir de marcadores microssatélites (DAVID *et al.*, 2018). Os progenitores da população MDP são originários de quatro países diferentes, sendo 13 de Uganda, uma introdução do Quênia, do Peru, e dos EUA cada, tendo sido selecionados com base em atributos desejáveis para um ideótipo de batata-doce, incluindo conteúdo de matéria seca, rendimento de raízes de armazenamento e teor de  $\beta$ -caroteno (Tabela 2).

Para realização dos ensaios KASP, realizou-se a coleta de amostras foliares em uma placa de 96 poços de 1,2 mL, consistindo em 94 poços com amostras (dois discos foliares) mais dois poços mantidos sem amostras para servirem de controle negativo das reações (ou “brancos”). As 94 amostras representaram material foliar de 93 indivíduos (Tabela 3), incluindo, 86 amostras do painel, sendo 81 progênies MDP e 5 dos seus genitores (‘Ejumula’, ‘Magabali’, ‘Mugande’, ‘SPK004’ e ‘NASPOT 11’). Adicionalmente, outros 7 acessos

Tabela 2 – Descrição dos genitores da população *Mwanga Diversity Panel* provenientes de duas populações denominadas A e B quanto ao país de origem e cor da polpa

Nome do acesso	Código do campo	Origem	Cor da polpa
<b>Genitores masculinos (população A)</b>			
Ejumula	A1	Uganda	Amarela pálida
NASPOT 1	A2	Uganda	Amarela escura
Dimbuka Bukulula	A3	Uganda	Creme escura
NASPOT 5/58	A4	Uganda	Creme escura
NASPOT 7	A5	Uganda	Laranja intermediária
SPK004	A6	Quênia	Amarela pálida
NASPOT 10 O	A7	Uganda	Laranja escura
NK259L	A8	Uganda	Creme escura
<b>Genitores femininos (população B)</b>			
Resisto	B1	EUA	Laranja escura
Magabali	B2	Uganda	Creme escura
NASPOT 5	B3	Uganda	Amarela escura
Wagabolige	B4	Uganda	Branca
Mugande	B5	Uganda	Creme escura
NASPOT 11	B6	Uganda	Creme escura
New Kawogo	B7	Uganda	Creme escura
Huarmeyano	B8	Peru	Amarela pálida

Fonte: Adaptado de Mugisa et al. (2022)

(‘NASPOT 8’, ‘Beauregard’, ‘Tanzania’, ‘Cemsa 74-228’, ‘Silk Omuyaka’, ‘D20’ e ‘NAROSPOT 1’, sendo o último presente em duplicata) foram amostrados. Os discos foliares foram desidratados em sílica gel por 48h, e enviados para extração de DNA e obtenção das intensidades alélicas para os alelos dos SNPs via KASP pela empresa Intertek (Suíça).

### 2.3. Dados genotípicos

Inicialmente, os dados brutos de intensidade dos alelos  $A$  e  $B$  foram inspecionados por meio de gráficos de dispersão e os indivíduos observados de acordo com a classificação fornecida pelo relatório da empresa em homozigóticos (fixados para um dos dois alelos  $A$  ou  $B$ ) ou heterozigóticos (uma mistura de alelos  $A$  e  $B$  em alguma proporção possível para indivíduos hexaploides), mas combinados em uma única classe, numa chamada pseudo-diploide. Essa análise inicial permitiu também acessar os valores de intensidade presentes nos controles negativos (‘brancos’) e identificar eventuais falhas de amplificação (designados pela empresa como ‘não classificados’). As falhas foram caracterizadas pelos baixos sinais de intensidade, próximos àqueles dos controles negativos, ou que eram insuficientes para permitir classificação em homozigoto ou heterozigoto.

Em seguida, procedeu-se com a dosagem alélica por meio do pacote R *fitPoly* v3.0.0 (VAN GEEST *et al.*, 2017). O algoritmo do *fitPoly*, inicialmente descrito para espécies tetraploides e implementado no pacote R *fitTetra* (VOORRIPS; GORT; VOSMAN, 2011), utiliza as razões de intensidade dos alelos do SNP dadas por  $X_i = B_i / (A_i + B_i)$ , em que  $A_i$  e  $B_i$  são as intensidades dos respectivos alelos  $A$  e  $B$  do SNP para o indivíduo  $i$ . Com base na distribuição da razão de sinal dos alelos, até sete normais podem ser ajustadas, cada uma correspondendo a uma classe de dosagem (de zero a seis cópias do alelo), constituindo, portanto, uma mistura de normais (VOORRIPS; GORT; VOSMAN, 2011).

Primeiro, uma função automática, denominada *saveMarkerModels* foi utilizada com os seguintes valores de parâmetros: *maxiter* = 0, *p.threshold* = 0.6, *peak.threshold* = 1, *call.threshold* = 0. O objetivo desses novos valores foi evitar que marcadores fossem descartados antes mesmo de terem uma chamada realizada, já que há vários níveis de restringências passados como padrão. Eventuais ajustes dos parâmetros das normais, isto é, médias determinadas para cada SNP, e desvios-padrão fixo igual a 0.02, foram realizados usando a função *CodomMarker* a fim de testar o ajuste dos modelos de mistura com base em uma inspeção gráfica de histogramas exibindo a frequência dos indivíduos (eixo  $y$ ) ao longo das razões  $X_i$  (eixo  $x$ ).

Tabela 3 – Pedigree dos 93 genótipos de batata-doce genotipados para validação dos marcadores KASPs. ‘NAROSPOT 1’ foi amostrado em duplicata, e ‘NA’ indica informação indisponível

Nome do acesso	Genitor feminino	Genitor masculino	Nome do acesso	Genitor feminino	Genitor masculino
UGP20170672-26	Wagabolige	SPK004	UGP20170493-30	NASPOT 5	Dimbuka Bukulula
UGP20170490-985	NASPOT 5	NASPOT 7	UGP20170342-23	Magabali	NK259L
UGP20170350-1	Magabali	Dimbuka Bukulula	UGP20170673-173	Wagabolige	NK259L
UGP20170347-11	Magabali	NASPOT 7	UGP20170024-9	Resisto	NK259L
UGP20170672-28	Wagabolige	SPK004	UGP20170907-17	NASPOT 11	NASPOT 7
UGP20170907-11	NASPOT 11	NASPOT 7	UGP20170907-31	NASPOT 11	NASPOT 7
UGP20170802-18	Mugande	Dimbuka Bukulula	UGP20170350-20	Magabali	Dimbuka Bukulula
UGP20170032-16	Resisto	Dimbuka Bukulula	UGP20170029-22	Resisto	NASPOT 10 O
UGP20170476-16	NASPOT 5	NASPOT 5/58	UGP20170029-15	Resisto	NASPOT 10 O
UGP20170907-10	NASPOT 11	NASPOT 7	UGP20170902-63	NASPOT 11	SPK004
UGP20170490-982	NASPOT 5	NASPOT 7	UGP20170679-13	Wagabolige	Dimbuka Bukulula
UGP20170943-39	New Kawogo	NASPOT 7	UGP20170802-11	Mugande	Dimbuka Bukulula
UGP20170334-27	Magabali	Ejumula	UGP20170783-12	Mugande	NASPOT5/58
UGP20170024-2	Resisto	NK259L	UGP20170342-22	Magabali	NK259L
UGP20171155-17	Huarmeyano	Dimbuka Bukulula	UGP20170902-56	NASPOT 11	SPK004
UGP20171141-10	Huarmeyano	Ejumula	UGP20171148-10	Huarmeyano	SPK004
UGP20170341-37	Magabali	SPK004	UGP20170342-15	Magabali	NK259L
UGP20170902-52	NASPOT 11	SPK004	UGP20170490-989	NASPOT 5	NASPOT 7
UGP20170028-1	Resisto	NASPOT 7	UGP20170342-17	Magabali	NK259L
UGP20170028-7	Resisto	NASPOT 7	UGP20170342-18	Magabali	NK259L
UGP20170934-42	New Kawogo	SPK004	UGP20171143-11	Huarmeyano	NASPOT5/58
UGP20170342-21	Magabali	NK259L	UGP20170934-49	New Kawogo	SPK004
UGP20170903-101	NASPOT 11	NK259L	UGP20170783-17	Mugande	NASPOT5/58
UGP20170910-18	NASPOT 11	Dimbuka Bukulula	UGP20170475-22	NASPOT 5	Ejumula
UGP20170342-12	Magabali	NK259L	UGP20170893-15	NASPOT 11	Ejumula
UGP20170944-17	New Kawogo	NASPOT 10 O	UGP20170350-18	Magabali	Dimbuka Bukulula
UGP20170893-11	NASPOT 11	Ejumula	UGP20171155-13	Huarmeyano	Dimbuka Bukulula
UGP20170342-14	Magabali	NK259L	UGP20170947-14	New Kawogo	Dimbuka Bukulula
UGP20170334-13	Magabali	Ejumula	UGP20170348-1	Magabali	NASPOT 10 O
UGP20170486-162	NASPOT 5	SPK004	UGP20170024-13	Resisto	NK259L
UGP20170793-120	Mugande	NK259L	UGP20170024-12	Resisto	NK259L
UGP20170674-16	Wagabolige	NASPOT 1	UGP20170939-14	New Kawogo	NASPOT 1
UGP20170015-26	Resisto	NASPOT 5/58	UGP20170893-13	NASPOT 11	Ejumula
UGP20170674-11	Wagabolige	NASPOT 1	UGP20170935-19	New Kawogo	NK259L
UGP20170674-13	Wagabolige	NASPOT 1	Magabali	NA	NA
UGP20170893-20	NASPOT_11	Ejumula	SPK004	NA	NA
UGP20170341-20	Magabali	SPK004	Mugande	NA	NA
UGP20170902-54	NASPOT 11	SPK004	NASPOT 11	NA	NA
UGP20170679-18	Wagabolige	Dimbuka Bukulula	Ejumula	NA	NA
UGP20170350-10	Magabali	Dimbuka Bukulula	Tanzania	NA	NA
UGP20170335-6	Magabali	NASPOT5/58	Silk Omuyaka	NA	NA
UGP20171152-1	Huarmeyano	NASPOT 7	Beauregard	NA	NA
UGP20170342-24	Magabali	NK259L	Cemsa 74-228	NA	NA
UGP20170342-32	Magabali	NK259L	NAROSPOT 1	NA	NA
UGP20170674-15	Wagabolige	NASPOT 1	NASPOT 8	NA	NA
UGP20170944-8	New Kawogo	NASPOT 10 O	D20	NA	NA
UGP20170342-19	Magabali	NK259L			

Fonte: *International Potato Center* (CIP).

## 2.4. Dados fenotípicos

### 2.4.1. Conteúdo de matéria seca, cor da polpa e cor da casca

Progênes MDP e seus genitores foram avaliados para os caracteres conteúdo de matéria seca, e cores da polpa e da casca em experimentos conduzidos em campo em Uganda nos anos de 2016 e 2017 pelo *National Crops Resources Research Institute* (NaCRRI), Namulonge, no âmbito do projeto *Genomic Tools for Sweetpotato Improvement* (GT4SP) (MUGISA *et al.*, 2022). O delineamento experimental do tipo linha-coluna (46 linhas  $\times$  43 colunas) consistiu na aleatorização de 1.896 progênes, 16 genótipos parentais e duas testemunhas ('New Kawogo' e 'Ejumula') replicadas em cada coluna (MUGISA *et al.*, 2022). O mesmo delineamento foi reproduzido em duas estações experimentais ao longo de dois anos: 2019A e 2019B em Abi Zonal Agricultural Research and Development Institute (Abi ZARDI), em Arua district, e 2019B e 2020A em Ngetta ZARDI, em Lira district, Uganda. Não houve análise via equipamento de espectrometria do infravermelho próximo (do inglês, *near-infrared spectrometry* – NIRS) e, portanto, não há informação sobre teores de amido, sacarose, glicose ou  $\beta$ -caroteno, tampouco teor de  $\beta$ -amilase foi obtido.

Após a colheita, a porcentagem de conteúdo de matéria seca de cada clone foi obtida cortando-se em pequenos pedaços duas raízes frescas de armazenamento de tamanho médio (aproximadamente 200-300 g) selecionadas aleatoriamente. Uma subamostra composta de cerca de 100-200 g foi então obtida, colocada em um saco de papel rotulado e pesada para determinar o peso fresco exato. A subamostra foi então seca a 70 °C em estufa por 72 h até massa constante ser atingida. A massa seca foi posteriormente pesada e o peso foi expresso como uma porcentagem do peso fresco (MUGISA *et al.*, 2022).

A análise dos dados foi realizada usando o pacote ASReml-R v. 4.1 no *software* R v. 4.0 (R DEVELOPMENT CORE TEAM, 2019). Um modelo misto conjunto para cada conteúdo de matéria seca foi usado da seguinte forma:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \mathbf{T}\mathbf{i} + \mathbf{e}_{\xi\eta}$$

em que  $\mathbf{y}$  refere-se ao vetor de observações fenotípicas,  $\boldsymbol{\beta}$  é o vetor de efeitos fixos de ambientes, e  $\mathbf{X}$  é a matriz delineamento;  $\mathbf{g}$  é o vetor aleatório de efeitos dos genótipos,  $\mathbf{g} \sim N(\mathbf{0}, \mathbf{A}\sigma_g^2)$ , em que  $\mathbf{A}$  é uma matriz de parentesco baseada em pedigree entre genótipos e  $\sigma_g^2$  é a variância genética aditiva, e  $\mathbf{Z}$  é a matriz de incidência;  $\mathbf{i}$  é o vetor aleatório de interação genótipo  $\times$  ambiente,  $\mathbf{i} \sim N(\mathbf{0}, \mathbf{I}\sigma_i^2)$ , em que  $\mathbf{I}$  é uma matriz identidade e  $\sigma_i^2$  é a variância da interação genótipo-por-ambiente, e  $\mathbf{T}$  é a matriz de delineamento;  $\mathbf{e}_{\xi\eta}$  é o vetor aleatório de resíduos, modelado usando um processo auto-regressivo de ordem 1 (AR1  $\times$  AR1) nas direções

de linha e coluna dentro de cada ambiente. A matriz de parentesco baseada em pedigree (**A**) foi calculado utilizando a função *Amatrix* do pacote R AGHmatrix v. 2.0.4 (AMADEU *et al.*, 2016) considerando autohexaploidia (MUGISA *et al.*, 2022).

Nos mesmos experimentos, a cor da polpa foi observada visualmente atribuindo-se notas de 1 a 9, em que: 1- branca; 2- creme clara; 3- creme escura; 4- amarela pálida; 5- amarela escura; 6- laranja pálida; 7- laranja escura; 8- laranja intermediária, e 9- laranja escura. Para a cor da casca foram atribuídas, também, notas de 1 a 9, em que: 1-branca; 2- creme; 3- amarela; 4- laranja; 5- marrom alaranjada; 6- rosada; 7- vermelha; 8- roxa avermelhada, e 9- roxa escura. Para ambos os fenótipos, médias foram obtidas ao longo dos ambientes para cada um dos genótipos.

#### 2.4.2. Firmeza da raiz

O caráter firmeza foi utilizado para avaliar os marcadores associados a teor de  $\beta$ -amilase em função da importante correlação negativa ( $-0.40$ ) entre esses caracteres (NAKITTO *et al.*, 2022) e à ausência do fenótipo teor de  $\beta$ -amilase propriamente dito. Neste trabalho anterior, genótipos foram submetidos à avaliação sensorial, com base em notas, quanto à firmeza e consistência ao toque e ao paladar. Os resultados da avaliação sensorial apresentaram correlação positiva com massa seca, absorção de água e tempo ótimo de cozimento das raízes. Um modelo para predição de firmeza foi desenvolvido considerando esses caracteres e medições mecânicas obtidas a partir de um analisador de textura conforme descrito a seguir.

Para a obtenção de variáveis, pedaços de raiz cozida foram analisados usando um analisador de textura, texturômetro TA-XT (Plus Stable Macro Systems, Godalming, Reino Unido) com célula de carga de 10 kg, seguindo um procedimento de análise de perfil de textura (NAKITTO *et al.*, 2022). Neste método, uma sonda de placa de compressão de 60 mm de diâmetro movendo-se a uma velocidade de 100 mm/min comprimiu a amostra (altura vertical de 2 cm) por 5 s (distância de 8,35 mm), descansando por 5 s entre as compressões, e a firmeza sendo determinada como o pico de força positiva (NAKITTO *et al.*, 2022). Os valores de medição obtidos para cada genótipo foram aplicados ao modelo de predição de firmeza, e os resultados obtidos foram utilizados nas análises de regressão.

### 2.5. Regressão linear simples e múltipla

Inicialmente, procurou-se ajustar os efeitos dos SNPs sob a variação dos caracteres fenotípicos em questão utilizando modelo de regressão linear simples:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

em que  $Y_i$  é o fenótipo do indivíduo  $i$ ,  $\beta_0$  é o intercepto,  $\beta_1$  é o efeito do SNP, e  $\varepsilon_i$  é o resíduo considerando  $\varepsilon_i \sim N(0, \sigma^2)$ . A informação do SNP para o indivíduo  $i$ ,  $X_i$ , foi inicialmente codificada de três formas a partir da informação de dosagem alélica obtida utilizando-se o fitPoly:

- Modelo aditivo em que:

$$X_i = \begin{cases} 0 & \text{se } AAAAAA \\ 1 & \text{se } AAAAAAB \\ 2 & \text{se } AAAAABB \\ 3 & \text{se } AAABBBB \\ 4 & \text{se } AABBBBB \\ 5 & \text{se } ABBBBBB \\ 6 & \text{se } BBBBBBB \end{cases}$$

- Modelo dominante para o alelo de referência  $A$  ('Dom. REF') em que:

$$X_i = \begin{cases} 0 & \text{se } BBBBBB \\ 1 & \text{se } A\_ \_ \_ \_ \_ \end{cases}$$

- Modelo dominante para o alelo alternativo  $B$  ('Dom. ALT') em que:

$$X_i = \begin{cases} 0 & \text{se } AAAAAA \\ 1 & \text{se } B\_ \_ \_ \_ \_ \end{cases}$$

Adicionalmente, utilizou-se um modelo quantitativo, que não depende da dosagem alélica pois foi simplesmente baseado na razão de intensidade alélica. Assim:

$$X_i = B_i / (A_i + B_i)$$

Os SNPs mais significativos foram escolhidos via teste de hipóteses utilizando a estatística  $F$  ( $P < 0,001$ ). Esse teste teve como objetivo verificar se a média do fenótipo dos indivíduos variava linearmente em função do genótipo do SNP de acordo com o modelo escolhido (aditivo ou dominante) ou em função da razão entre os sinais dos alelos. As hipóteses associadas ao teste foram:

- $H_0: \beta_1 = 0$  (a média do fenótipo não varia linearmente em função da codificação do SNP, ou seja, o efeito do SNP não é significativo).
- $H_1: \beta_1 \neq 0$  (a média do fenótipo varia linearmente em função da codificação do SNP, ou seja, o efeito do SNP é significativo).

Finalmente, considerando as codificações aditivas e dominantes para os marcadores, procedeu-se com seleção *stepwise*, conforme implementada na função *step* do pacote do R MASS v.7.3-60. Nesta função, o modelo inicial que só continha apenas o intercepto foi sendo incrementado, a cada passo, com marcadores significativos (busca *forward*). Em seguida, as variáveis são novamente testadas, e eventualmente excluídas do modelo se, marginalmente às demais variáveis, não for mais significativa (eliminação *backward*). A decisão sobre entrada e

saída de variáveis foi realizada pelo Critério de Informação Bayesiana (do inglês, *Bayesian information criterion – BIC*) (SCHWARZ, 1978), dado por:

$$BIC = -2 \ln(\mathcal{L}) + k \ln(N)$$

em que  $\mathcal{L}$  é a verossimilhança do modelo,  $k$  é o número de parâmetros e  $N$  é o tamanho da amostra.

Com esta análise foi possível verificar se uma regressão linear múltipla, combinando efeitos de mais de um marcador simultaneamente, possibilitava uma maior explicação da variação do caráter em questão. Marcadores com mais de 15% de informações faltantes foram eliminados e, em seguida, apenas o conjunto de fenótipo mais marcadores sem dados perdidos foram considerados pela função *step*. Finalmente, os SNPs selecionados foram ajustados conjuntamente em um modelo de regressão linear múltipla:

$$Y_i = \beta_0 + \sum_{m=1}^M \beta_m X_{im} + \varepsilon_i$$

em que  $\beta_m$  é o efeito do marcador  $m$  sob dada codificação, com  $m = 1, \dots, M$ , e  $M$  é o número de combinações marcador-codificação selecionadas.

### 3. RESULTADOS E DISCUSSÃO

#### 3.1. Desenvolvimento dos SNPs

De um total de 60 SNPs convertidos em KASPs, 8 foram para teor de amido/matéria seca, 12 para teor de  $\beta$ -caroteno/cor da polpa, 15 para teor de glicose, 10 para cor da casca, e 15 para teor de  $\beta$ -amilase (Tabela 4).

Para teor de amido/matéria seca e teor de  $\beta$ -caroteno/cor de polpa, picos colocalizados na região 3,10 Mpb do cromossomo 3 indicaram, previamente, a existência de dois genes ligados, sendo um envolvido na rota da síntese de amido, *itf03g05100*, anotado como *sucrose synthase (SuSY)*, e outro elencado na rota de produção de  $\beta$ -caroteno, *itf03g05110*, anotado como *phyotene synthase (PSY)*. Esses genes já tinham sido identificados como putativamente relacionados aos respectivos caracteres e, a partir do sequenciamento de RNA (RNA-seq), demonstraram padrões de expressão diferencial entre os genitores ‘Beauregard’ e ‘Tanzania’ (GEMENET *et al.*, 2020). Similarmente, o QTL para  $\beta$ -caroteno/cor de polpa na região 22,12 Mpb do cromossomo 12 já havia sido usado para identificação do gene *itf12g24270*, anotado como *Orange*, que funciona como um “interruptor” molecular relacionado à formação de amilocromoplastos, organelas nas quais carotenoides podem ser estavelmente acumulados (GEMENET *et al.*, 2020; OSORIO, 2019).

Para teor de glicose, dois genes identificados próximos ao QTL na região de 11,57 Mpb do cromossomo 3 estão implicados na síntese de açúcares. O primeiro, *itf03g13310*, anotado como *Glycoside hydrolase family 33 (GH33)*, codifica para uma enzima membro da família das glicosil hidrolases, conhecida pelo papel importante na montagem e desmontagem polissacarídeos de glicose presentes em parede celular vegetal, como a celulose, e de reserva, como o amido (KFOURY *et al.*, 2024). O segundo, *itf03g14590*, anotado como *Uridine diphospho (UDP)-glucose*, é um nucleotídeo-açúcar com participação-chave na biossíntese de sacarose em plantas vasculares (FIGUEROA; LUNN; IGLESIAS, 2021), um dissacarídeo constituído de glicose e frutose.

Para cor da casca, dois genes anotados como *Myeloblastosis (MYB) viral oncogene homolog*, *itf12g04070* e *itf12g04100*, foram encontrados na região 2,4 Mpb do QTL identificado no cromossomo 12. Essa classe de fatores de transcrição, uma das maiores em plantas, é caracterizada pela conservação do domínio homólogo a MYB e são implicados na regulação de biossíntese de metabólitos secundários, incluindo antocianinas (CAO *et al.*, 2020). Genes caracterizados como *MYB homolog* foram apontados como componentes-chave na

Tabela 4 – Informação sobre SNPs convertidos em KASPs. Anotação e posição referem-se ao genoma de *Ipomoea trifida* v.3

Caráter	Gene	Anotação do gene	Cromossomo	Posição (pb)	Nome
Teor de amido/ matéria seca	<i>itf03g05100</i>	<i>Sucrose synthase (SuSY)</i>	3	3102534	snpIB00061
				3102587	snpIB00062
				3103751	snpIB00063
				3104439	snpIB00064
				3104492	snpIB00065
				3104982	snpIB00066
				3105556	snpIB00067
				3105685	snpIB00068
				3120204	snpIB00069
				3120259	snpIB00070
Teor de β-caroteno/ cor da polpa	<i>itf03g05110</i>	<i>Phytene synthase (PSY)</i>	3	3120823	snpIB00071
				3120853	snpIB00072
				3120976	snpIB00073
				3121746	snpIB00074
				3121840	snpIB00075
				11402767	snpIB00076
				11402997	snpIB00077
				11403022	snpIB00078
				11403267	snpIB00079
				11403882	snpIB00080
Teor de glicose	<i>itf03g13310</i>	<i>Glycoside hydrolase family 33 (GH33)</i>	3	11404117	snpIB00081
				11404495	snpIB00082
				11404546	snpIB00083
				12113895	snpIB00084
				12114120	snpIB00085
	<i>itf03g14590</i>	<i>Uridine diphospho (UDP)-glucose</i>	3	12114210	snpIB00086
				12115371	snpIB00087
				12115625	snpIB00088
				12117047	snpIB00089
				12118347	snpIB00090
Cor da casca	<i>itf12g04070</i>	<i>Myeloblastosis (MYB) viral oncogene homolog</i>	12	2426382	snpIB00091
				2426724	snpIB00092
				2427587	snpIB00093
				2427638	snpIB00094
				2428136	snpIB00095
	<i>itf12g04100</i>	<i>Myeloblastosis (MYB) viral oncogene homolog</i>	12	2456039	snpIB00096
				2456362	snpIB00097
				2456605	snpIB00098
				2457504	snpIB00099
				2457510	snpIB00100
Teor de β-caroteno/ cor da polpa	<i>itf12g24270</i>	<i>Orange</i>	12	22122584	snpIB00101
				22122650	snpIB00102
				22122703	snpIB00103
				22123861	snpIB00104
				22124356	snpIB00105
Teor de β-amilase	<i>itf13g16360</i>	<i>β-amylase</i>	13	18655399	snpIB00106
				18655409	snpIB00107
				18655765	snpIB00108
				18655789	snpIB00109
				18655985	snpIB00110
				18656184	snpIB00111
				18656199	snpIB00112
				18656283	snpIB00113
				18656347	snpIB00114
				18656632	snpIB00115
				18656687	snpIB00116
				18657194	snpIB00117
				18657261	snpIB00118
18657330	snpIB00119				
18657547	snpIB00120				

Fonte: o autor.

coloração de casca vermelha de inúmeras espécies, como batata (ENDELMAN; JANSKY, 2016) e maçã (WANG *et al.*, 2020), por exemplo.

Para teor de  $\beta$ -amilase, o gene *itf13g16360*, anotado como  $\beta$ -*amylase*, foi identificado no QTL presente no cromossomo 12, na região em torno de 18,65 Mpb. Previamente a este trabalho, análises de expressão diferencial deste gene em batata-doce apontou sua relação com a textura da raiz cozida (BANDA *et al.*, 2021).  $\beta$ -amilase quebra as moléculas de amido por hidrólise, liberando maltose e maltotriose, que podem ser metabolizados para glicose e/ou serem reciclados para biossíntese de amido (SMITH; ZEEMAN; SMITH, 2005). Nas raízes tuberosas de batata-doce,  $\beta$ -amilase representa 5% das proteínas totais (NAKAMURA *et al.*, 1991), sendo responsável pelo aumento do teor de açúcares durante o cozimento da raiz ao liberar maltose pela quebra do amido (KUMAGAI *et al.*, 1990).

### 3.2. Dosagem alélica

Para cada SNP, a metodologia de genotipagem KASP foi utilizada, retornando uma medida de intensidade para cada alelo,  $A_i$  e  $B_i$ , por indivíduo  $i$ . No programa de melhoramento de CIP-Uganda, o desenvolvimento de marcadores KASP é terceirizado e, portanto, um trabalho mínimo foi necessário para a preparação da amostra. No entanto, apenas uma chamada pseudo-diploide foi fornecida pela empresa, na qual os indivíduos homozigóticos foram relativamente bem determinados (próximos aos eixos), mas os heterozigotos ficaram combinados em apenas uma classe (Figura 2). Espécies autohexaploides, como a batata-doce, podem apresentar até sete classes genóticas por loco (de 0 a 6), das quais até cinco correspondem a classes heterozigóticas, perdendo informação importante para análises genéticas posteriores. Sendo assim, para determinação da dosagem alélica hexaploide, fez-se uso de métodos mais apropriados.

A distribuição dos sinais de cada genótipo para cada marcador SNP, junto à classificação fornecida pela empresa (Figura 2) exibiu um padrão esperado para genotipagem pseudo-diploide, a única que foi disponibilizada. Isto porque os indivíduos heterozigóticos foram combinados em uma única classe, conforme mencionado, enquanto os indivíduos homozigóticos foram, em geral, bem classificados. No entanto, a expectativa de que indivíduos homozigóticos sempre estariam localizados nos extremos da distribuição da nuvem de pontos criou equívocos nas chamadas, como nos KASPs *snpIB00064* e *snpIB00078*, para os quais, na ausência de genótipos fixados para cada um dos alelos  $A$  e  $B$ , respectivamente, indivíduos heterozigóticos foram retornados como homozigóticos por estarem nas extremidades das

nuvens. Considerando todos os marcadores (um total de 96 amostras  $\times$  60 KASPs = 5760 pontos), 1510 pontos (26,2%) foram classificados como homozigóticos ( $AA = 0$  ou  $BB = 2$ ) e 3053 pontos (53,0%) foram classificados como heterozigóticos ( $AB = 1$ ) a partir da chamada pseudo-diploide fornecida pela empresa.

Outro importante detalhe desta primeira inspeção gráfica foi a existência de indivíduos não classificados, em poucas ocasiões para alguns marcadores (como snpIB00061 e snpIB00065), sobretudo quando os pontos se encontravam nos limites de classificação entre nuvens de pontos considerados homozigóticos e heterozigóticos. No entanto, alguns marcadores tiveram a totalidade de indivíduos não classificados. Foram os casos de snpIB00066, snpIB00089, snpIB00103, snpIB00104, e snpIB00105, aparentemente monomórficos, e snpIB00084 e snpIB00097, aparentemente polimórficos. O total de pontos não classificados foi 617 (10,7%), considerando todos os marcadores em conjunto.

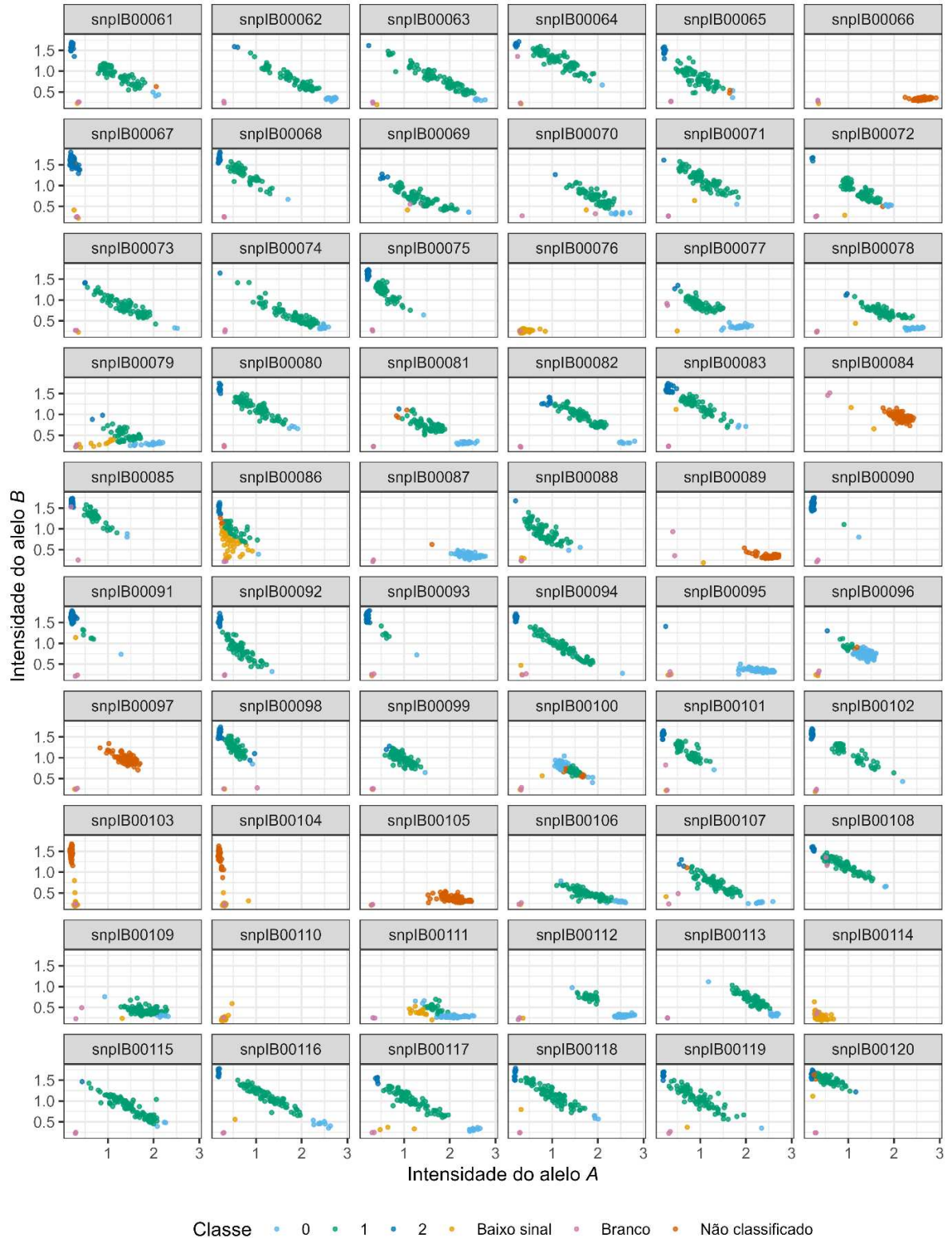
O total de amostras classificados como ‘branco’ foi 120 (dois poços por KASP). Amostras consideradas de baixo sinal, ou seja, que não alcançaram intensidade adequada para chamada de genótipos, conforme determinação do *software* proprietário da empresa, totalizaram 460 (8,0%). Baixo sinal foi encontrado para todas as amostras para os marcadores snpIB00076, snpIB00110 e snpIB00114, considerados como falhas de amplificação, portanto. Outros três marcadores, snpIB00079, snpIB00086, snpIB00111, apresentaram baixo sinal em várias amostras, quando comparados aos demais. Para classificação no fitPoly, foram excluídos pontos marcados como ‘brancos’ e baixo sinal, representando um total de 580 pontos (10,1%). Os resultados sugerem que os marcadores desenvolvidos apresentam diferentes taxas de eficiência de amplificação, o que pode ser observado ao comparar o acúmulo de sinal de cada marcador. Por um lado, snpIB00084 aparentemente possui alta eficiência de amplificação, refletindo, inclusive, no relativo acúmulo de sinal nas amostras denominadas ‘branco’ (isto é, reações realizadas na ausência de DNA molde). Por outro lado, snpIB00079 apresenta baixa eficiência de amplificação. O procedimento ideal por parte da empresa seria estabelecer os melhores parâmetros para o processo de genotipagem para cada marcador, ou estabelecer um grupo de marcadores que apresente o mesmo perfil de amplificação sob os mesmos parâmetros.

Uma vez que os sinais emitidos pelos alelos apresentavam geralmente um padrão contínuo, e não perfeitamente proporcional à dosagem, o pacote fitPoly foi utilizado para determinar a dosagem mais provável. Para tanto, inicialmente, a função automática *saveMarkerModels* foi utilizada (Figura 3). De modo geral, a chamada funcionou relativamente bem ao permitir a separação de classes heterozigóticas mais evidentes sob inspeção visual. No entanto, 23 marcadores (snpIB00066, 068, 070, 071, 072, 074, 085, 086, 088, 091, 093, 094,

099, 100, 103, 104, 106, 107, 108, 109, 111, 115, 119) ainda apresentaram problemas nas chamadas, como duas classes combinadas em uma única. Dessa maneira, procedeu-se com a utilização da função *CodomMarker* para todas as marcas (Figura 4). Na comparação dos resultados das duas funções, notou-se que a maioria dos marcadores apresentaram chamadas muito similares, se não idênticas. Esta função permitiu melhorar as chamadas dos marcadores mencionados anteriormente, exceto snpIB00070, snpIB00074 e snpIB00094. No caso desses, procedeu-se com ajustes manuais dos parâmetros de médias das normais, a partir das observações dos histogramas de frequência de indivíduos ao longo das razões  $X_i = B_i/(A_i + B_i)$ , também com auxílio da função *CodomMarker* (Figura 5).

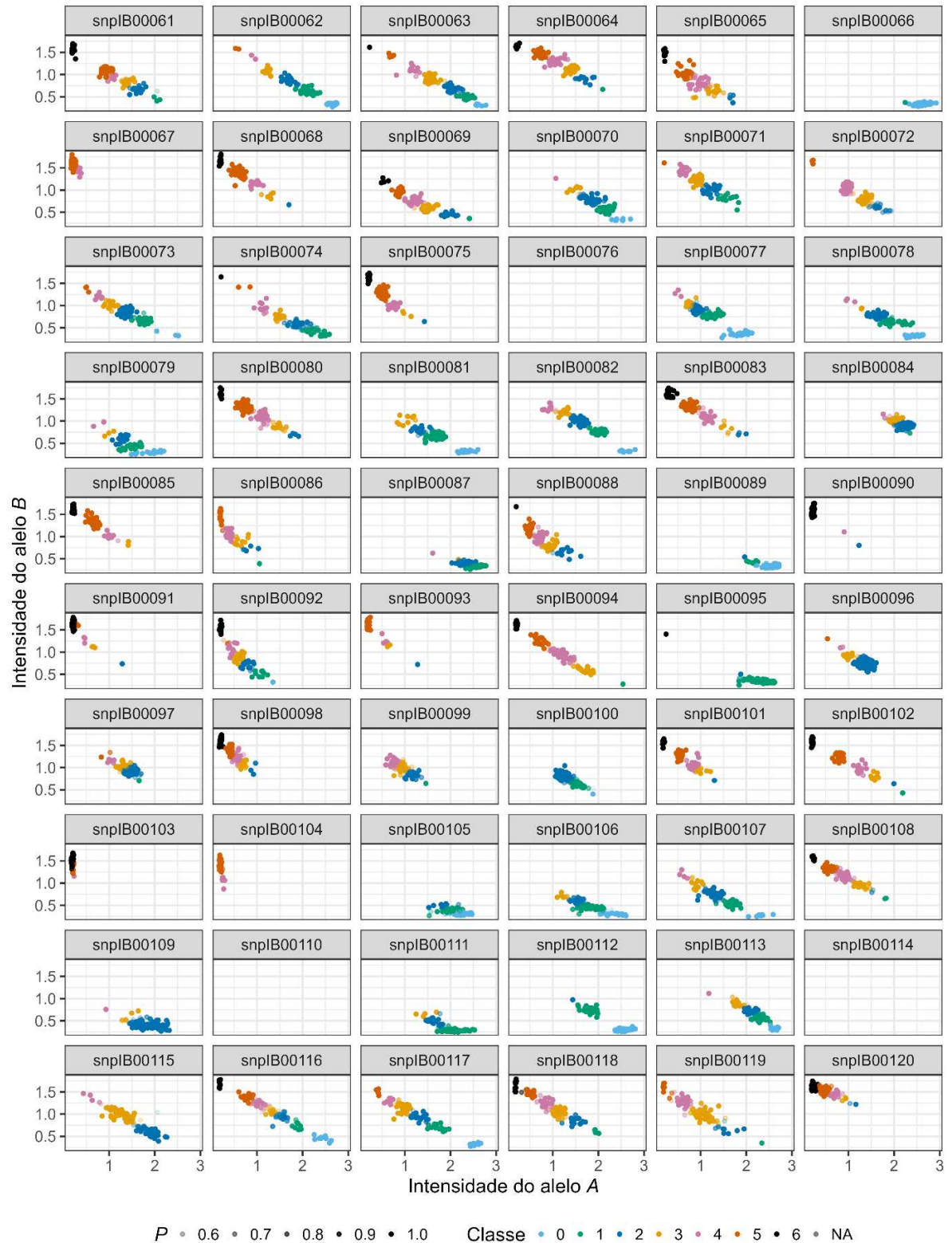
O conjunto de marcadores com necessidade de ajustes adicionais, envolvendo a função *CodomMarker* e entrada manual das médias das normais, pode variar de um conjunto de indivíduos para outro. Isso porque classes fenotípicas não representadas aqui podem ser representadas em outras populações, facilitando a definição das classes. Ainda que possível de ser realizada, essas correções manuais constituem uma atividade monótona e demorada, pois exige avaliação gráfica dos resultados de chamada a cada nova rodada. No futuro, à medida que novos indivíduos são genotipados com os mesmos marcadores, seria aconselhável comparar as razões ao longo das diferentes rodadas e, eventualmente, determinar médias para cada uma das normais, procurando padronizá-las para um conjunto de dados qualquer. Isso facilitaria a dosagem alélica e aceleraria as decisões a serem tomadas pelos melhoristas com auxílio dos KASPs.

Figura 2 – Gráfico de dispersão das intensidades de sinal dos alelos *A* e *B* para 60 KASPs. As cores exibem a classificação fornecida pela empresa que originou os dados. ‘Branco’ significa que não foi adicionada amostra de DNA à reação, enquanto ‘baixo sinal’ e ‘não classificado’ quer dizer que os pontos não receberam qualquer uma das três possíveis classes pseudo-diploides,  $AA = 0$ ,  $AB = 1$  ou  $BB = 2$



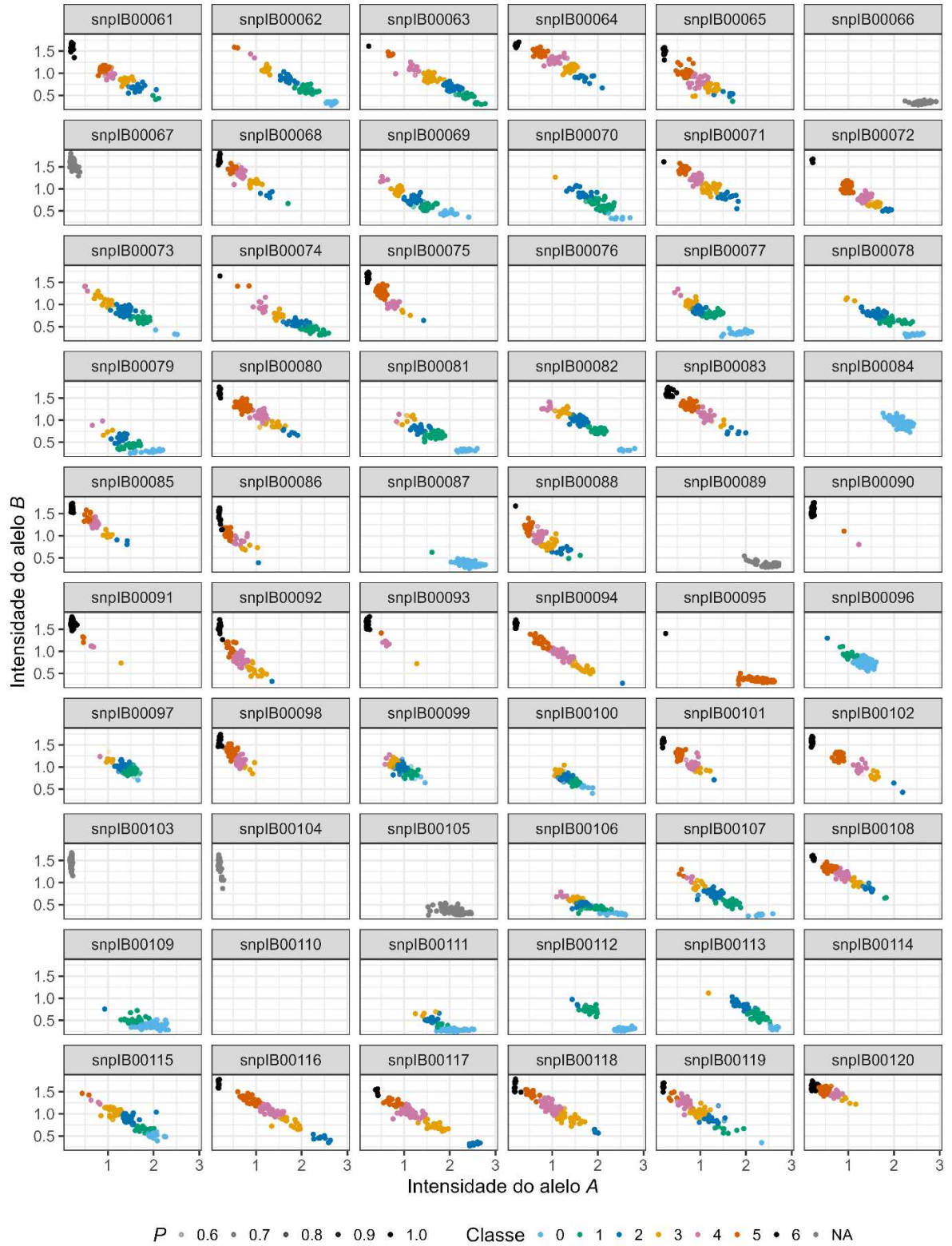
Fonte: o autor.

Figura 3 – Classificação obtida a partir da função *saveMarkerModels* do pacote do R *fitPoly*. As cores referem-se às doses mais prováveis enquanto a transparência refere-se ao valor de  $P$  associadas às doses. Os dados foram filtrados para eliminar os ‘brancos’ e as amostras contendo baixo sinal. Dados perdidos (‘NA’) resultaram de falha na chamada



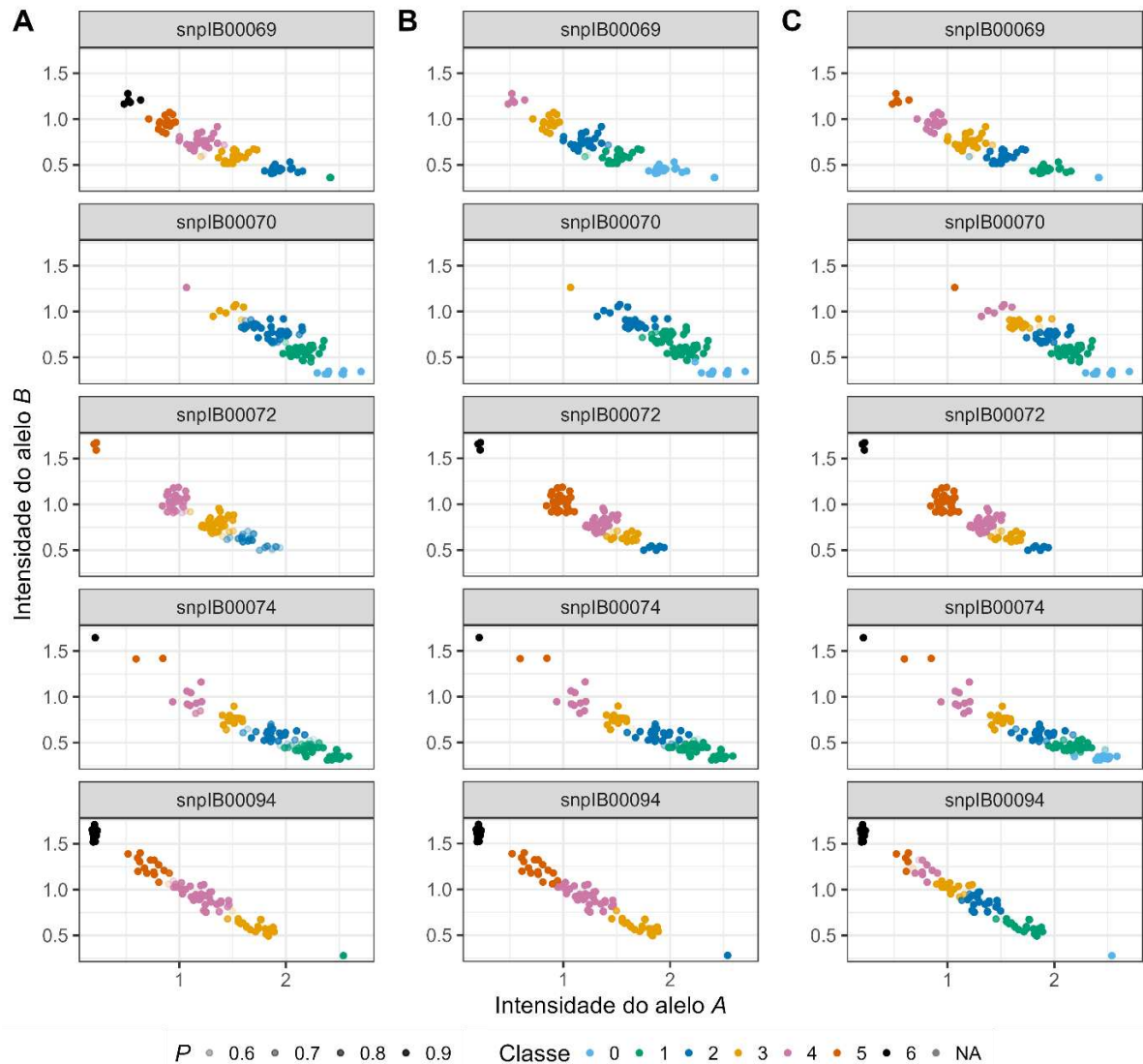
Fonte: o autor.

Figura 4 – Classificação obtida a partir da função *CodomMarker* do pacote do R *fitPoly* mais ajustes manuais. As cores referem-se às doses mais prováveis enquanto a transparência refere-se ao valor de  $P$  associadas às doses. Os dados foram filtrados para eliminar os ‘brancos’ e as amostras contendo baixo sinal. Dados perdidos (‘NA’) resultaram de falha na chamada



Fonte: o autor.

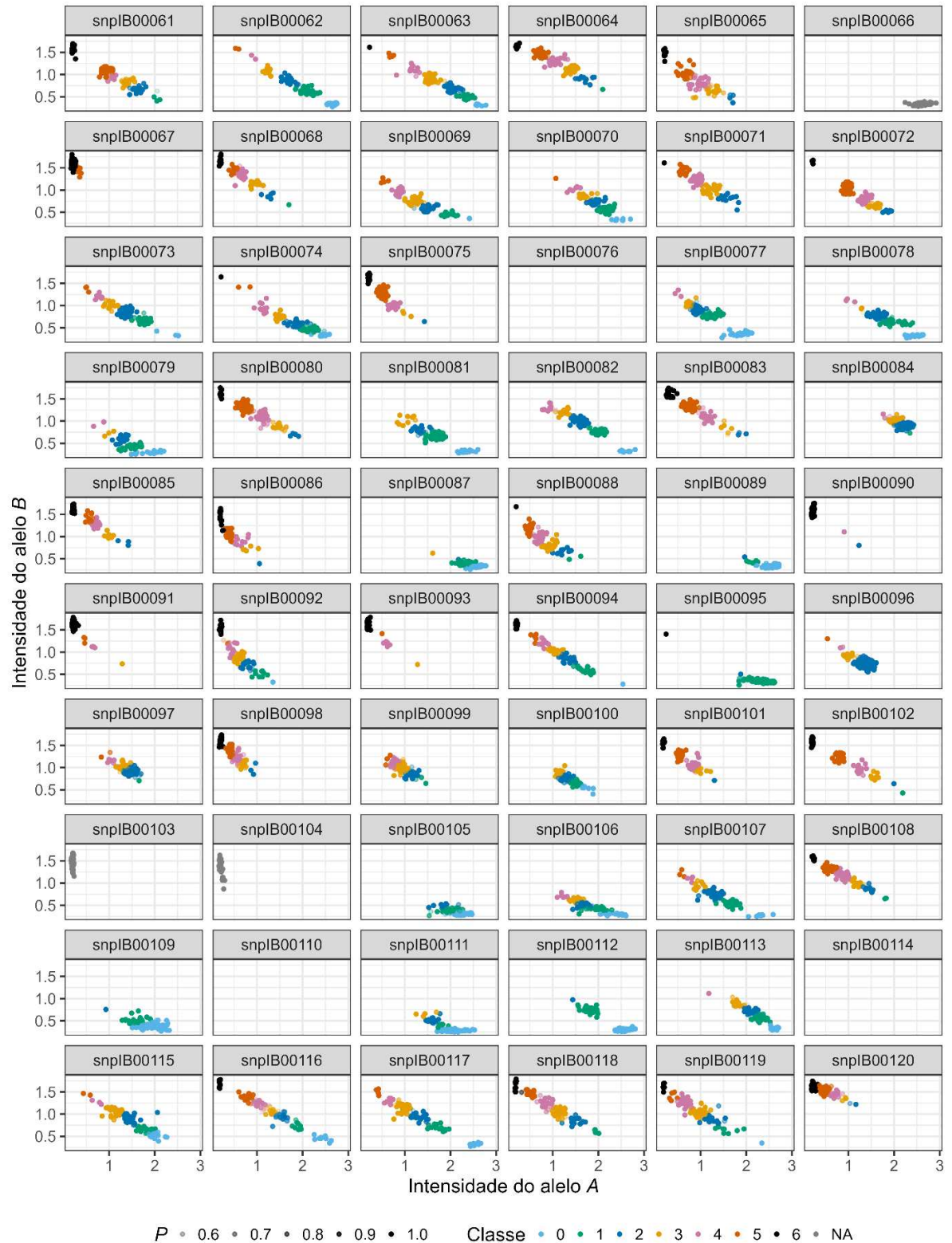
Figura 5 – Comparação de resultados da dosagem alélica usando funções do fitPoly *saveMarker Models* (A), *CodomMarker* (B) e correção manual (C). Os marcadores snpIB00069 e snpIB00072 apresentaram melhores resultados a partir das funções *saveMarkerModels* e *CodomMarker*, respectivamente. Os marcadores snpIB00070, snpIB00074 e snpIB00094 exigiram correção manual



Fonte: o autor.

Em resumo, dos 60 SNPs, três SNPs falharam em produzir sinal (snpIB00076, snpIB00110 e snpIB00114), e três SNPs foram considerados monomórficos (snpIB00066, snpIB00103 e snpIB00104). Além disso, apesar da existência de polimorfismos, outros três SNPs foram excluídos por apresentar uma frequência alélica mínima (do inglês, *minor allele frequency* – MAF) inferior a 10% (snpIB00087, snpIB00090 e snpIB00095). De um total de 4794 pontos presentes na chamada final dos 51 SNPs restantes (Figura 6), observaram-se 1244 indivíduos homozigóticos (26,0%), 3439 (71,7%) indivíduos heterozigóticos, 111 (2,3%) dados faltantes, incluindo ‘brancos’ e amostras contendo baixo sinal, notando-se uma melhora em relação aos 10,7% de dados faltantes originalmente entregues pela empresa.

Figura 6 – Classificação final em classes de doses alélicas obtidas a partir do fitPoly. As cores referem-se às doses mais prováveis enquanto a transparência refere-se ao valor de  $P$  associadas às doses. Os dados foram filtrados para eliminar os ‘brancos’ e as amostras contendo baixo sinal. Dados perdidos (‘NA’) resultaram de falha na chamada



Fonte: o autor.

### 3.2. Validação dos marcadores

Modelos de regressão linear simples entre os valores fenotípicos e as codificações variadas envolvendo dosagem ou razão das intensidades alélicas foram ajustados para cada um dos 51 SNPs polimórficos. Dos 93 indivíduos genotipados com KASP, 86 haviam sido fenotipados em campo e, portanto, este foi o número máximo de amostras utilizado nas regressões. De acordo com a estatística  $F$  ( $P < 0,001$ ), cinco, onze, sete e dois SNPs foram significativamente associados aos fenótipos de conteúdo de matéria seca, cor da polpa, cor da casca e firmeza, respectivamente (Tabela 5 e Figura 7).

Para conteúdo de matéria seca, três marcadores localizados no gene *itf03g05100* (*SuSY*), snpIB00061, snpIB00062 e snpIB00065, e um localizado no gene *itf03g05110* (*PSY*), snpIB00070, foram significativamente associados ao caráter. Adicionalmente, um SNP localizado no gene *itf03g13310* (*GH33*), snpIB00079, distante fisicamente do QTL-alvo, apresentou associação significativa. Já para cor de polpa, apenas os marcadores snpIB00065 e snpIB00067 no gene *itf03g05100* (*SuSY*), e snpIB00070 no gene *itf03g5110* (*PSY*), próximos ao QTL-alvo, foram significativamente associados ao caráter. Conteúdo de matéria seca (amido) e cor de polpa (teor de  $\beta$ -caroteno) são caracteres negativamente correlacionadas (GEMENET *et al.*, 2020), o que, além da proximidade física entre os genes envolvidos, pode explicar a associação significativa de SNPs do gene não diretamente relacionado com a expressão do caráter.

No caso do marcador snpIB00079, presente no gene *GH33* potencialmente envolvido com teor de glicose, a associação positiva com teor de matéria seca pode ser devida ao balanço entre a quebra do amido e a produção de glicose. A hidrólise do amido libera maltose, que é formado por duas moléculas de glicose, e o metabolismo da maltose libera glicose nos tecidos (SMITH; ZEEMAN; SMITH, 2005). A manipulação genética de enzimas envolvidas no metabolismo de carbono em plantas também demonstrou que existe um balanço entre as quantidades de amido e glicose. Por exemplo, a superexpressão do regulador metabólico *sucrose non-fermenting-1-related protein kinase-1* (*SnRK1*) em batatas levou a um decréscimo significativo no teor de glicose (17%~56%) e a um aumento no teor de amido (23%~30%) nos tubérculos (MCKIBBIN *et al.*, 2006), enquanto a expressão heteróloga de uma *vacuolar invertase* de batata-doce, *Ib $\beta$ fruct2*, diminuiu o conteúdo de amido e aumentou o teor de glicose em folhas de *Arabidopsis* (ZHANG *et al.*, 2023). Desta maneira, a associação significativa com conteúdo de matéria seca de um marcador intuitivamente relacionado com teor de glicose pode ser devido ao *crosstalk* metabólico entre amido e glicose.

Interessantemente, todos os sete marcadores snpIB00077~083 do gene *itf03g13310* (*GH33*) e o marcador snpIB00098 do gene *itf12g04100* (*MYB homolog*) também exibiram sinal significativo de associação para cor de polpa, apesar da distância em relação ao QTL-alvo. Não surpreendentemente, a glicose também está envolvida na biossíntese de carotenoides. Em plantas, os carotenoides são principalmente produzidos nos plastídios a partir dos substratos gerados pela via MEP (2-C-methyl-D-erythritol 4-phosphate). Gliceraldeído-3-fosfato e piruvato são os substratos iniciais da via MEP, e ambos os substratos são disponibilizados via glicólise (CAZZONELLI; POGSON, 2010; MOISE; AL-BABILI; WURTZEL, 2014). Considerando que a grande maioria dos SNPs selecionados no gene *GH33* apresentaram associação positiva com a cor da polpa, há a possibilidade de que os teores de glicose nas raízes da batata-doce exercem grande influência na concentração de carotenoides na polpa através da regulação da disponibilidade de substratos para a biossíntese dos mesmos.

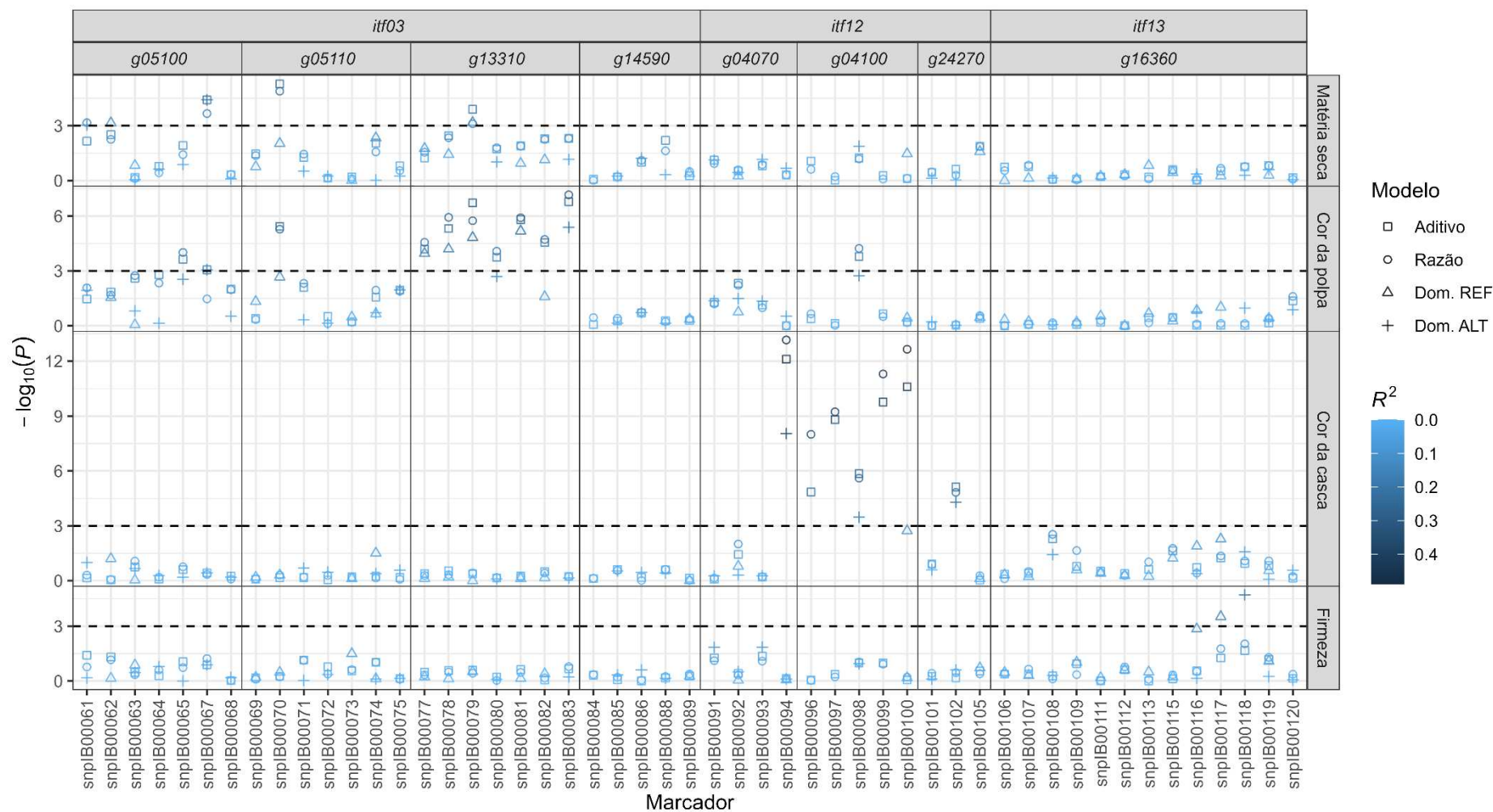
Para cor de casca, o marcador snpIB00094 do gene *itf12g04070* (*MYB homolog*) e todos os cinco marcadores snpIB00096~100 do gene *itf12g04100* (*MYB homolog*) exibiram significância. Além deles, snpIB00102 do gene *itf12g24270* (*Orange*) apresentou sinal significativo de associação com o caráter. Finalmente, para firmeza, apenas dois marcadores, snpIB00117 e snpIB00118 presentes no gene *itf13g16360* ( *$\beta$ -amylase*), reportaram associação significativa com o caráter. O único gene cujos SNPs não apresentaram associação com qualquer caráter avaliado foi *itf03g14590* (*UDP-glucose*) no QTL para teor de glicose. Marcadores presentes na mesma região e em elevado desequilíbrio de ligação (Figura 8) mostraram associações significativas para os mesmos caracteres, notadamente todos pertencentes ao gene *itf03g13310*, e a maioria dos SNPs contidos nos genes *itf03g05100*, *itf03g05110*, *itf12g04100* e *itf13g16360*.

Tabela 5 – Marcadores KASPs com efeitos significativos ( $P < 0,001$ ) para os caracteres conteúdo de matéria seca, cor de polpa, cor de casca e firmeza para diferentes modelos de codificação da informação do SNP, e seus respectivos interceptos e coeficiente de determinação ( $R^2$ )

Caráter	Marcador	Codificação	Intercepto	Efeito	$P$	$R^2$	
Matéria seca	snpIB00061	Razão	35,4237	-2,9214	6,80e-04	0,13	
		Dom. ALT	32,7515	1,4938	9,08e-04	0,12	
	snpIB00062	Dom. REF	32,7532	1,4920	6,93e-04	0,13	
		snpIB00067	Aditivo	17,1667	2,8207	3,86e-05	0,19
	Razão		8,9089	28,3843	2,15e-04	0,15	
	snpIB00070	Dom. ALT	34,0909	-2,8207	3,86e-05	0,19	
		Aditivo	35,1489	-0,7269	5,18e-06	0,22	
	snpIB00079	Razão	36,3361	-9,6354	1,30e-05	0,20	
		Aditivo	34,7180	-0,8030	1,26e-04	0,18	
		Razão	35,5010	-7,1648	7,69e-04	0,14	
		Dom. REF	34,9053	-1,5227	6,69e-04	0,14	
	Cor da polpa	snpIB00065	Aditivo	1,9926	0,5439	2,33e-04	0,15
Razão			2,5623	3,3674	9,74e-05	0,16	
snpIB00067		Aditivo	15,7959	-1,9413	8,71e-04	0,13	
		Dom. ALT	4,1480	1,9413	8,71e-04	0,13	
snpIB00070		Aditivo	3,2041	0,6234	3,68e-06	0,22	
		Razão	2,1330	8,4660	5,29e-06	0,22	
snpIB00077		Aditivo	3,5186	0,5616	6,40e-05	0,17	
		Razão	2,7118	4,0437	2,73e-05	0,19	
		Dom. REF	3,3079	1,3608	1,11e-04	0,16	
snpIB00078		Aditivo	3,3931	0,7137	4,81e-06	0,22	
		Razão	2,4541	7,3442	1,18e-06	0,25	
		Dom. REF	3,2343	1,4290	6,38e-05	0,17	
snpIB00079		Aditivo	3,3513	0,8916	1,90e-07	0,30	
		Razão	2,3923	8,3336	1,81e-06	0,26	
		Dom. REF	3,1968	1,6064	1,48e-05	0,22	
snpIB00080		Aditivo	6,8944	-0,5909	1,81e-04	0,15	
		Razão	6,5003	-3,8430	8,39e-05	0,17	
snpIB00081		Aditivo	3,3811	0,8181	1,57e-06	0,24	
		Razão	2,5111	6,3895	1,24e-06	0,24	
		Dom. REF	3,1281	1,5838	6,55e-06	0,21	
snpIB00082		Aditivo	3,1476	0,6153	2,81e-05	0,19	
		Razão	2,3324	5,4001	1,90e-05	0,19	
snpIB00083		Aditivo	8,2100	-0,8112	1,65e-07	0,28	
		Razão	7,5508	-5,1712	6,91e-08	0,29	
		Dom. ALT	3,0683	1,6488	4,14e-06	0,22	
snpIB00098		Aditivo	6,7365	-0,5271	1,63e-04	0,16	
		Razão	8,7051	-5,8952	5,90e-05	0,18	
		Cor da casca	snpIB00094	Aditivo	8,0204	-0,8465	7,74e-13
Razão				9,0790	-7,1472	6,92e-14	0,49
Dom. ALT				2,7603	3,1947	8,99e-09	0,33
snpIB00096	Aditivo		9,4612	-1,8630	1,41e-05	0,20	
	Razão		12,3132	-18,3843	9,96e-09	0,33	
snpIB00097	Aditivo		9,7664	-1,8282	1,56e-09	0,35	
	Razão		15,5357	-25,0772	5,87e-10	0,37	
snpIB00098	Aditivo		0,8575	0,9626	1,40e-06	0,24	
	Razão		-2,1276	9,9509	2,48e-06	0,23	
	Dom. ALT		6,6361	-1,8506	3,39e-04	0,14	
snpIB00099	Aditivo		0,6729	1,5442	1,72e-10	0,39	
	Razão		-4,7236	19,5749	4,99e-12	0,44	
snpIB00100	Aditivo	2,5051	1,6692	2,52e-11	0,42		
	Razão	-3,5922	25,1553	2,23e-13	0,48		
snpIB00102	Aditivo	0,4284	0,9655	7,50e-06	0,21		
	Razão	2,0756	4,8167	1,48e-05	0,20		
	Dom. ALT	6,3798	-1,8925	5,13e-05	0,18		
Firmeza	snpIB00117	Dom. REF	6,0379	-1,5328	2,99e-04	0,15	
	snpIB00118	Dom. ALT	6,1464	-1,6991	1,94e-05	0,20	

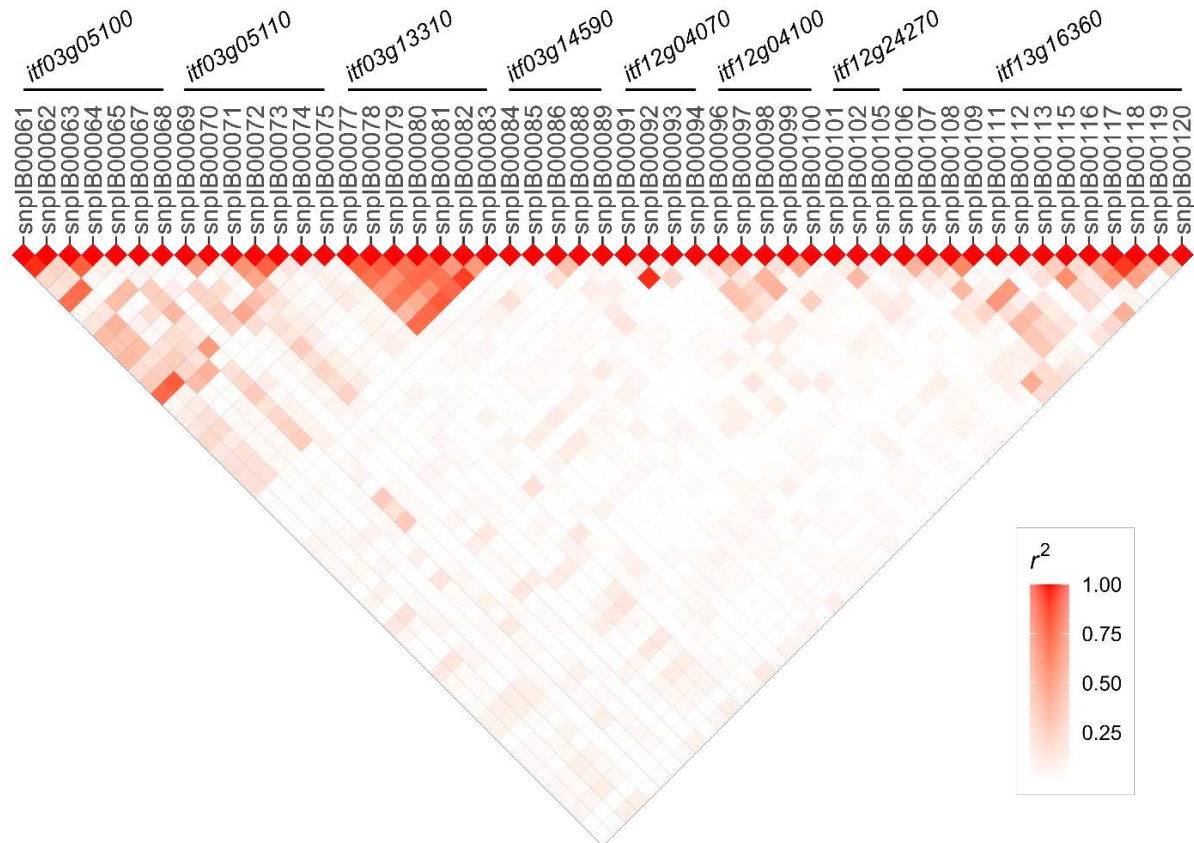
Fonte: o autor.

Figura 7 – Valores de  $P$ , transformados usando  $-\log_{10}$ , para os 51 KASPs testados contra quatro caracteres em população diversa de batata-doce ( $N = 86$ ) considerando quatro modelos de codificação dos genótipos. Os coeficientes de determinação  $R^2$  a partir de modelos de regressão simples são expressos em gradiente de cor. Efeitos considerados significativos ( $P < 0,001$ ) ficam acima da linha tracejada. Os KASPs estão agrupados de acordo com os genes anotados no genoma de *Ipomoea trifida* v.3, a partir dos quais SNPs-alvo foram selecionados



Fonte: o autor.

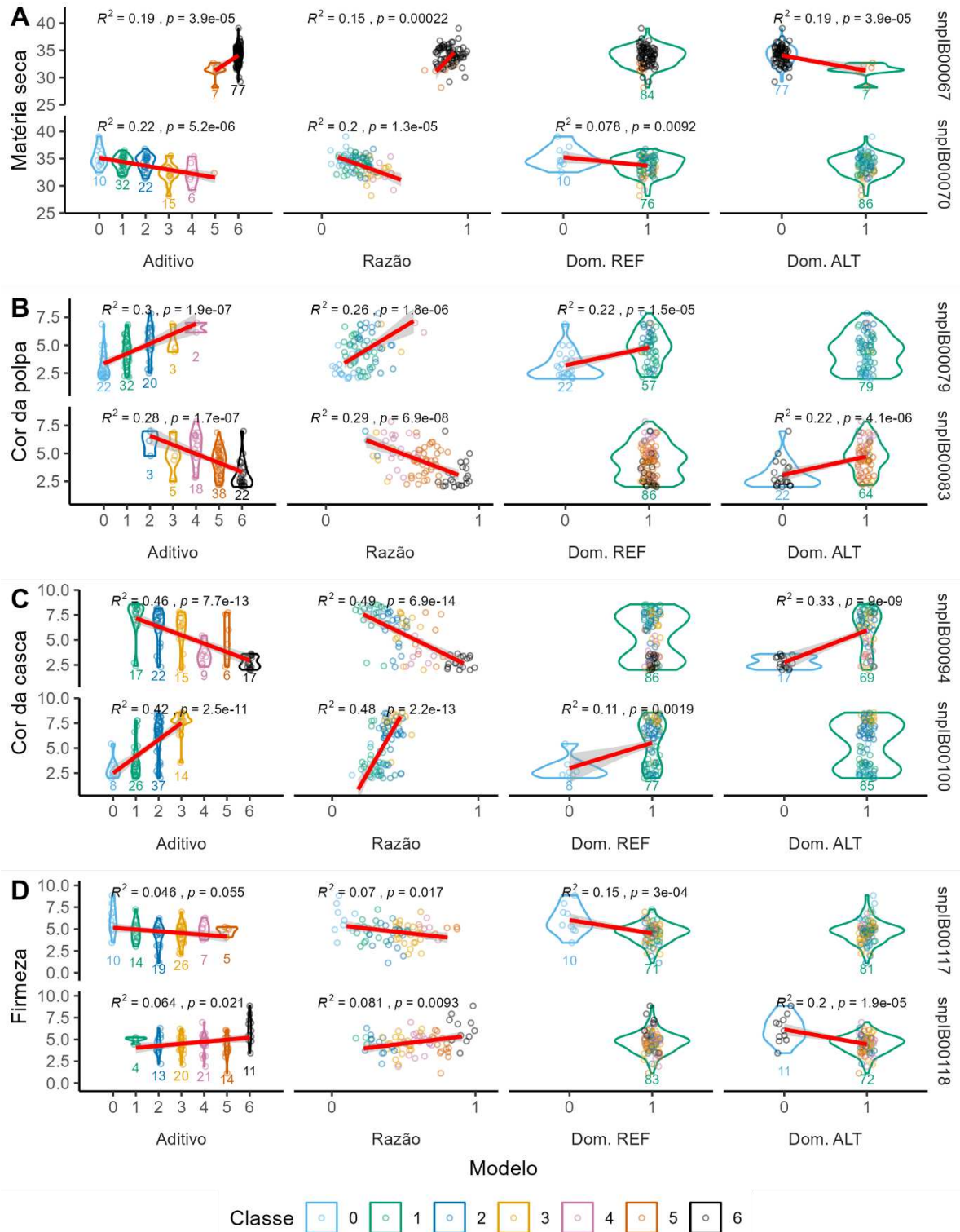
Figura 8 – Desequilíbrio de ligação a partir das correlações ao quadrado ( $r^2$ ) entre as dosagens alélicas de 51 marcadores KASP genotipados em clones de batata-doce ( $N = 93$ )



Fonte: o autor.

As associações mais significativas foram: snpIB00067 ( $P = 3,9 \times 10^{-5}$ ,  $R^2 = 0,19$  para os modelos ‘Aditivo’ e ‘Dom. ALT), e snpIB00070 ( $5,2 \times 10^{-6}$ ,  $R^2 = 0,22$  para o modelo ‘Aditivo’) para conteúdo de matéria seca (Figura 9A); snpIB00079 ( $1,9 \times 10^{-7}$ ,  $R^2 = 0,30$  para o modelo ‘Aditivo’) e snpIB00083 ( $6,8 \times 10^{-8}$ ,  $R^2 = 0,29$  para o modelo ‘Razão’) para cor de polpa (Figura 9B); snpIB00094 ( $6,9 \times 10^{-14}$ ,  $R^2 = 0,49$  para o modelo ‘Razão’) e snpIB00100 ( $2,2 \times 10^{-13}$ ,  $R^2 = 0,48$  para o modelo ‘Aditivo’) para cor de casca (Figura 9C); e snpIB00117 ( $3,0 \times 10^{-4}$ ,  $R^2 = 0,15$  para o modelo ‘Dom. REF) e snpIB00118 ( $1,9 \times 10^{-5}$ ,  $R^2 = 0,20$  para o modelo ‘Dom. ALT) para firmeza (Figura 9D). De modo geral, os modelos baseados na razão das intensidades alélicas (‘Razão’), foram semelhantes em performance para detectar associação quando comparados aos modelos aditivos para os SNPs significativamente associados a conteúdo de matéria seca, cor de polpa e cor de casca (Figura 7 e Figura 8A-C). Isso é interessante do ponto de vista prático, pois não haveria necessidade de realizar chamada de doses. Por outro lado, se o controle genético do caráter for o de dominância, a chamada de doses ainda se fez necessária para detectar associações – e realizar seleção assistida –, como no caso dos SNPs significativos para firmeza (Figura 7 e Figura 9D).

Figura 9 – Regressão linear considerando quatro modelos de codificação dos genótipos para os dois SNPs com efeitos mais significativos para cada caráter: matéria seca (A), cor da polpa (B), cor da casca (C), e firmeza (D). Nos modelos ‘Dom. REF’ e ‘Dom. ALT’, os genótipos homocigóticos para os respectivos alelos *A* e *B* (classe 0) foram testados contra o conjunto dos demais genótipos (classe 1). Os números de indivíduos são mostrados sob cada classe



Fonte: o autor.

Amido é o maior constituinte da matéria seca em batata-doce (WALTER Jr *et al.*, 2000), sendo, portanto, teor de amido e conteúdo de matéria seca dois caracteres positivamente correlacionados, com valores de correlação variando entre 0,77 e 0,89 (CERVANTES-FLORES *et al.*, 2011; GEMENET *et al.*, 2020; YADA *et al.*, 2017). Não surpreendentemente, QTLs colocalizados foram identificados para ambos os caracteres (GEMENET *et al.*, 2020). Já teor de  $\beta$ -amilase e firmeza são dois caracteres negativamente correlacionados, com valor de correlação  $-0,40$  (BANDA *et al.*, 2021). O amido é um dos grandes contribuintes para a firmeza da polpa. Seu conteúdo e composição afetam o resultado final do cozimento com relação à maciez e textura (ALLAN *et al.*, 2024; BANDA *et al.*, 2021; YOON *et al.*, 2018), o que coloca a  $\beta$ -amilase em um ponto central como proteína influenciadora na característica, ainda que não completamente determinante.

Tanto conteúdo de matéria seca como firmeza são caracteres que possivelmente abrangem outros constituintes, uma vez que fibras, açúcares mais simples e proteínas estão presentes nas raízes (DE ALBUQUERQUE; SAMPAIO; DE SOUZA, 2019). Diferença entre os tipos de amidos também podem explicar variação na firmeza (ALLAN *et al.*, 2024). Desse modo, ainda que conteúdo de matéria seca e firmeza possam estar relacionados a marcadores em desequilíbrio de ligação com QTLs mapeados para teor de amido e  $\beta$ -amilase, respectivamente, seria improvável conseguir explicar toda a variação dos caracteres usados neste estudo a partir de marcadores fortemente ligados a QTL para outros caracteres. Neste trabalho, não foi possível estudar a associação dos marcadores em relação teor de amido ou  $\beta$ -amilase em função da indisponibilidade dos mesmos.

Já teor de  $\beta$ -caroteno e cor de polpa são caracteres altamente correlacionados em batata-doce, com valores de correlação próximos a 0,90 (GEMENET *et al.*, 2020). Ou seja, na ausência de segregação para teor de antocianina, a cor da polpa é majoritariamente determinada pelo teor de  $\beta$ -caroteno. Já no caso da cor da casca, pouco se sabia até o momento sobre correlação com outros caracteres. Ao que parece, carotenoides e, sobretudo, antocianina acumuladas na casca são responsáveis pela explicação na variação de cor de casca em batata-doce (DRAPAL *et al.*, 2019).

No caso dos SNPs significativos para caracteres não relacionados ao QTL-alvo, em vez de interpretar eventuais associações como artefatos originados em função de problemas relacionados a amostragem, por exemplo, é possível estabelecer algumas hipóteses biológicas para explicar tais associações. SNPs relacionados a região originalmente designada ao controle para teor de glicose, presentes no gene *GH33 (itf03g13310)* e que apresentaram associação tanto para matéria seca (snpIB00079) como para cor da polpa (snpIB00077~083), podem ser

entendidos como capturando efeitos da partição de carbono – ora para carboidratos simples ou complexos ora para carotenoides –, comum a todos esses caracteres. Para o caso do gene *Orange* (*itf12g24270*), previamente mapeado para cor de polpa, pode-se especular que o marcador snpIB00102, significativo para cor de casca, esteja indicando que  $\beta$ -caroteno produzido e armazenado na polpa contribui, também, para variação observada na cor da casca. A ausência de prévia identificação de QTLs nessas regiões pode ser explicada pela menor diversidade nas populações de mapeamento (biparentais) empregadas nos estudos anteriores quando comparada à diversidade da população de validação (multiparental) utilizada neste estudo.

Investigando se regressão linear múltipla, ou seja, a combinação da informação de mais de um marcador testado sob codificações aditivas e dominantes, representaria um acréscimo significativo em termos de explicação da variação para cada caráter, procedeu-se com a seleção *stepwise* das variáveis. A análise retornou seis, cinco, seis e dois marcadores como variáveis explicativas da variação para os caracteres conteúdo de matéria seca, cor de polpa, cor de casca e firmeza, respectivamente (Tabela 6). A etapa de seleção *stepwise* considerou *BIC* como critério para seleção de marcadores, e quando os modelos foram ajustados com a função *lm* do R, a significância poderia entrar em desacordo com essa seleção, como, de fato, foi observado. Além disso, o tamanho amostral pode ter afetado a seleção das variáveis. Para a etapa de seleção, restringiu-se ao conjunto de *todos os marcadores* com dados completos, após eliminação de daqueles com mais de 15% de dados perdidos. Já para o ajuste final usando *lm*, apenas os dados perdidos para os *marcadores selecionados* foram eliminados.

Para conteúdo de matéria seca e firmeza, marcadores sob codificação aditiva (snpIB00070, 067, 105) foram selecionados, além daqueles sob codificação dominante (snpIB00119, 107, 100), sendo snpIB00100 e snpIB00105 não-significativos ( $P > 0,05$ ) após ajuste final do modelo de regressão linear múltipla. Já para cor de polpa (snpIB00079, 068, 071, 098, 085) e cor de casca (snpIB00099, 094, 097, 102, 065, 078), apenas marcadores com codificação aditiva foram selecionados, sendo snpIB00099 não-significativo ( $P > 0,05$ ) após ajuste final do modelo. No caso de firmeza, apenas o marcador com codificação dominante (snpIB00117) foi significativo quando modelado conjuntamente com um de codificação aditiva (snpIB00101), não significativo ( $P > 0,05$ ). Os respectivos valores de  $R^2$  múltiplo (e  $R^2_{aj}$  múltiplo ajustado) foram 0,46 (e 0,42) para conteúdo de matéria seca, 0,52 (e 0,48) para cor de polpa, 0,73 (e 0,71) para cor de casca, e 0,17 (e 0,15) para firmeza. Ou seja, houve acréscimo importante da explicação da variação total pelos marcadores para todos os caracteres exceto

Tabela 6 – Estimativas dos interceptos, efeitos e valores de  $P$  associados com marcadores KASP combinados em modelos de regressão linear múltipla para cada caráter, e seus respectivos coeficientes de determinação múltiplo ( $R^2$ ) e múltiplo ajustado ( $R^2_{aj}$ ). Os SNPs foram selecionados a partir de procedimento *stepwise* utilizando critério de informação Bayesiano e a significância foi expressa como \*\*\* $P < 0,001$ , \*\* $P < 0,01$  e \* $P < 0,05$

Caráter	Marcador	Codificação	Intercepto	Efeito	$P$	$R^2$	$R^2_{aj}$
Matéria seca	snpIB00070	Aditivo	23,5422	-0,6040	3,52e-05***	0,46	0,42
	snpIB00067	Aditivo		2,0393	1,04e-03**		
	snpIB00119	Dom. ALT		2,1829	3,21e-03**		
	snpIB00107	Dom. REF		-1,8437	8,02e-03**		
	snpIB00105	Aditivo		-0,4502	6,20e-02		
	snpIB00100	Dom. REF		-0,8307	1,19e-01		
Cor de polpa	snpIB00079	Aditivo	6,0324	0,6061	1,55e-04***	0,52	0,48
	snpIB00068	Aditivo		0,5089	2,47e-05***		
	snpIB00071	Aditivo		-0,5293	1,18e-03**		
	snpIB00098	Aditivo		-0,2817	1,74e-02*		
	snpIB00085	Aditivo		-0,2754	2,06e-02*		
Cor de casca	snpIB00099	Aditivo	4,3415	0,2906	1,89e-01	0,73	0,71
	snpIB00094	Aditivo		-0,5421	1,42e-07***		
	snpIB00097	Aditivo		-1,0327	3,50e-05***		
	snpIB00102	Aditivo		0,4933	4,43e-04***		
	snpIB00065	Aditivo		0,5564	1,23e-04***		
	snpIB00078	Aditivo		-0,4019	9,09e-03**		
Firmeza	snpIB00117	Dom. REF	5,1826	-1,6380	1,58e-04***	0,17	0,15
	snpIB00101	Aditivo		0,2036	1,60e-01		

Fonte: o autor.

firmeza. Esse resultado implicaria, portanto, na necessidade de genotipar mais de um único SNP para seleção dos caracteres conteúdo de matéria seca, cor da polpa e cor da casca.

Esse acréscimo na explicação da variação total faz sentido no contexto de espécies autoploidoides como a batata-doce, já que a informação bialélica do SNP pode ser relativamente limitada quando comparada à variação multialélica que um gene segregando na progênie pode exibir. Para aplicações em SAM, idealmente, um dos alelos dos SNPs deveria marcar inequivocadamente o(s) haplótipo(s) responsável(is) pela característica desejada. Ainda que isso possa ser eventualmente alcançado em espécies diploides com alguma facilidade, o mesmo pode não ser verdadeiro em espécies poliploides. Em estudos futuros, uma forma de direcionar alelos de SNP aos haplótipos específicos seria estudar a variação existente em populações multiparentais, em vez de biparentais, como frequentemente ocorreu nos estudos de descoberta de QTL que originaram as regiões-alvo neste trabalho.

#### 4. CONCLUSÕES

O objetivo básico no melhoramento é o desenvolvimento de cultivares de plantas mais produtivas com características desejáveis, constituindo um processo trabalhoso e demorado. A utilização de marcadores muito próximos a genes controlando caracteres de interesse permite acelerar a seleção de características desejadas, sobretudo oligogênicas e de alta herdabilidade, já que apenas DNA é necessário para a tomada de decisão (COLLARD; MACKILL, 2007). O uso de marcadores SNP na seleção assistida por marcadores permite que os melhoristas pré-selecionem genótipos que combinem características desejadas para subsequente avaliação de campo e rastreiem populações segregantes rapidamente, otimizando tempo e recursos quando da fenotipagem de caracteres de produção (COLLARD; MACKILL, 2007). Em batata-doce, espera-se que a SAM auxilie na redução de custos e aumente a eficácia e a precisão na seleção de cultivares promissoras na substituição de variedades locais que não contemplem características importantes para segurança alimentar e nutricional. De fato, os marcadores validados podem ser usados para selecionar as características do estudo simultaneamente e espera-se que abordem os desafios associados à deficiência de vitamina A e à maior procura por variedades com maior teor de matéria seca.

Neste trabalho, foi possível validar um conjunto de marcadores KASP que podem ser aplicados na SAM, facilitando o desenvolvimento de novas cultivares de batata-doce. Os SNPs selecionados aqui são promissores por (i) estarem associados a genes putativos relacionados aos caracteres de interesse, (ii) terem sido obtidos a partir de mapas de ligação densos, (iii) terem sido utilizados métodos para obtenção das dosagens, e (iv) terem sido testados contra painel com diversidade superior e diferente daquelas populações das quais foram obtidos. No entanto, eventuais dificuldades na obtenção das doses e necessidade de ajustes manuais podem dificultar a tomada de decisão na imediatez que um programa de melhoramento necessita, e isto precisa ser revisitado com novas metodologias de obtenção das doses ou aprimoramento daquelas existentes. Os marcadores com efeito aditivo, para os quais a razão entre os sinais dos alelos pode ser empregada, facilitará sua utilização pelos melhoristas, uma vez que a etapa de dosagem alélica pode ser eliminada. No entanto, essa abordagem não captura associações no caso de controle genético dominante, como, possivelmente, no caso da firmeza. Os SNPs com maiores associações explicam parcial e significativamente a variação fenotípica de caracteres de interesse, e podem ser utilizadas para a aplicação imediata da SAM para o melhoramento genético desses caracteres.

## REFERÊNCIAS

- ALLAN, M. C.; JOHANNINGSMEIER, S. D.; NAKITTO, M.; GUAMBE, O.; ABUGU, M.; PECOTA, K. V.; CRAIG YENCHO, G. Baked sweetpotato textures and sweetness: An investigation into relationships between physicochemical and cooked attributes. **Food Chemistry**, v. 21, p. 101072, 30 mar. 2024. <https://doi.org/10.1016/j.fochx.2023.101072>.
- AMANKWAAH, V. A. **Phenotyping and genetic studies of storage root chemistry traits in sweetpotato**. 2019. 300 f. Ph.D. Dissertation – North Carolina State University, Raleigh, 2019. Disponível em: <http://www.lib.ncsu.edu/resolver/1840.20/37131>.
- ANDRADE, M. I.; RICARDO, J.; NAICO, A.; ALVARO, A.; MAKUNDE, G. S.; LOW, J.; ORTIZ, R.; GRÜNEBERG, W. J. Release of orange-fleshed sweetpotato (*Ipomoea batatas* [L.] Lam.) cultivars in Mozambique through an accelerated breeding scheme. **The Journal of Agricultural Science**, v. 155, n. 6, p. 919–929, ago. 2017. <https://doi.org/10.1017/S002185961600099X>.
- BANDA, L.; KYALLO, M.; DOMELEVO ENTFELLNER, J.-B.; MOYO, M.; SWANCKAERT, J.; MWANGA, R. O. M.; ONYANGO, A.; MAGIRI, E.; GEMENET, D. C.; YAO, N.; PELLE, R.; MUZHINGI, T. Analysis of  $\beta$ -amylase gene (*Amy $\beta$* ) variation reveals allele association with low enzyme activity and increased firmness in cooked sweetpotato (*Ipomoea batatas*) from East Africa. **Journal of Agriculture and Food Research**, v. 4, p. 100121, 1 jun. 2021. <https://doi.org/10.1016/j.jafr.2021.100121>.
- BOHAR, R.; CHITKINENI, A.; VARSHNEY, R. K. Genetic Molecular Markers to Accelerate Genetic Gains in Crops. **BioTechniques**, v. 69, n. 3, p. 158–160, 1 set. 2020. <https://doi.org/10.2144/btn-2020-0066>.
- BOVELL-BENJAMIN, A. C. Sweet Potato: A Review of its Past, Present, and Future Role in Human Nutrition. **Advances in Food and Nutrition Research**. [S. l.]: Academic Press, 2007. v. 52, p. 1–59. DOI 10.1016/S1043-4526(06)52001-7. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1043452606520017>. Acesso em: 19 set. 2024.
- CAO, Y.; LI, K.; LI, Y.; ZHAO, X.; WANG, L. MYB Transcription Factors as Regulators of Secondary Metabolism in Plants. **Biology**, v. 9, n. 3, p. 61, mar. 2020. <https://doi.org/10.3390/biology9030061>.
- CAZZONELLI, C. I.; POGSON, B. J. Source to sink: regulation of carotenoid biosynthesis in plants. **Trends in Plant Science**, v. 15, n. 5, p. 266–274, 1 maio 2010. <https://doi.org/10.1016/j.tplants.2010.02.003>.
- CERVANTES-FLORES, J. C.; SOSINSKI, B.; PECOTA, K. V.; MWANGA, R. O. M.; CATIGNANI, G. L.; TRUONG, V. D.; WATKINS, R. H.; ULMER, M. R.; YENCHO, G. C. Identification of quantitative trait loci for dry-matter, starch, and  $\beta$ -carotene content in sweetpotato. **Molecular Breeding**, v. 28, p. 201–216, 2011. <https://doi.org/10.1007/s11032-010-9474-5>.
- COLLARD, B. C. Y.; MACKILL, D. J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. **Philosophical Transactions of the Royal Society**

**B: Biological Sciences**, v. 363, n. 1491, p. 557–572, 22 ago. 2007.  
<https://doi.org/10.1098/rstb.2007.2170>.

DA SILVA PEREIRA, G.; GEMENET, D. C.; MOLLINARI, M.; OLUKOLU, B. A.; WOOD, J. C.; DIAZ, F.; MOSQUERA, V.; GRÜNEBERG, W. J.; KHAN, A.; BUELL, C. R.; YENCHO, G. C.; ZENG, Z.-B. Multiple QTL mapping in autopolyploids: A random-effect model approach with application in a hexaploid sweetpotato full-sib population. **Genetics**, v. 215, n. 3, p. 579–595, jul. 2020. <https://doi.org/10.1534/genetics.120.303080>.

DAVID, M. C.; DIAZ, F. C.; MWANGA, R. O. M.; TUMWEGAMIRE, S.; MANSILLA, R. C.; GRÜNEBERG, W. J. Gene pool subdivision of East African sweetpotato parental material. **Crop Science**, v. 58, n. 6, p. 2302–2314, 2018.  
<https://doi.org/10.2135/cropsci2017.11.0695>.

DE ALBUQUERQUE, T. M. R.; SAMPAIO, K. B.; DE SOUZA, E. L. Sweet potato roots: Unrevealing an old food as a source of health promoting bioactive compounds – A review. **Trends in Food Science & Technology**, v. 85, p. 277–286, 1 mar. 2019.  
<https://doi.org/10.1016/j.tifs.2018.11.006>.

DRAPAL, M.; ROSSEL, G.; HEIDER, B.; FRASER, P. D. Metabolic diversity in sweet potato (*Ipomoea batatas*, Lam.) leaves and storage roots. **Horticulture Research**, v. 6, n. 1, p. 2, dez. 2019. <https://doi.org/10.1038/s41438-018-0075-5>.

ENDELMAN, J. B.; JANSKY, S. H. Genetic mapping with an inbred line-derived F2 population in potato. **Theoretical and Applied Genetics**, v. 129, n. 5, p. 935–943, 1 maio 2016. <https://doi.org/10.1007/s00122-016-2673-7>.

FAO, “Food and Agriculture Organization”. **Crops and livestock products**. 2023. **FAOSTAT**. Disponível em: <https://www.fao.org/faostat/en/>.

FIGUEROA, C. M.; LUNN, J. E.; IGLESIAS, A. A. Nucleotide-sugar metabolism in plants: the legacy of Luis F. Leloir. **Journal of Experimental Botany**, v. 72, n. 11, p. 4053–4067, 18 maio 2021. <https://doi.org/10.1093/jxb/erab109>.

GEMENET, D. C.; DA SILVA PEREIRA, G.; DE BOECK, B.; WOOD, J. C.; MOLLINARI, M.; OLUKOLU, B. A.; DIAZ, F.; MOSQUERA, V.; SSALI, R. T.; DAVID, M.; KITAVI, M. N.; BURGOS, G.; FELDE, T. Z.; GHISLAIN, M.; CAREY, E.; SWANCKAERT, J.; COIN, L. J. M.; FEI, Z.; HAMILTON, J. P.; ... BUELL, C. R. Quantitative trait loci and differential gene expression analyses reveal the genetic basis for negatively associated  $\beta$ -carotene and starch content in hexaploid sweetpotato [*Ipomoea batatas* (L.) Lam.]. **Theoretical and Applied Genetics**, v. 133, n. 1, p. 23–36, 1 jan. 2020.  
<https://doi.org/10.1007/s00122-019-03437-7>.

GRÜNEBERG, W. J.; DE BOECK, B.; DIAZ, F.; EYZAGUIRRE, R.; LOW, J. W.; REIF, J. C.; CAMPOS, H. Heterosis and responses to selection in orange-fleshed sweetpotato (*Ipomoea batatas* L.) improved using reciprocal recurrent selection. **Frontiers in Plant Science**, v. 13, 2022. DOI 10.3389/fpls.2022.793904. Disponível em: <https://www.frontiersin.org/articles/10.3389/fpls.2022.793904>. Acesso em: 4 jan. 2023.

IBGE, “Instituto Brasileiro de Geografia e Estatística”. PAM - Produção Agrícola Municipal. 2023. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9117-producao-agricola-municipal-culturas-temporarias-e-permanentes.html>.

- KATAYAMA, K.; KOBAYASHI, A.; SAKAI, T.; KURANOUCI, T.; KAI, Y. Recent progress in sweetpotato breeding and cultivars for diverse applications in Japan. **Breeding Science**, v. 67, n. 1, p. 3–14, 2017. <https://doi.org/10.1270/jsbbs.16129>.
- KFOURY, B.; RODRIGUES, W. F. C.; KIM, S.-J.; BRANDIZZI, F.; DEL-BEM, L.-E. Multiple horizontal gene transfer events have shaped plant glycosyl hydrolase diversity and function. **New Phytologist**, v. 242, n. 2, p. 809–824, 2024. <https://doi.org/10.1111/nph.19595>.
- KUMAGAI, T.; UMEMURA, Y.; BABA, T.; IWANAGA, M. The inheritance of  $\beta$ -amylase null in storage roots of sweet potato, *Ipomoea batatas* (L.) Lam. **Theoretical and Applied Genetics**, v. 79, n. 3, p. 369–376, 1 maio 1990. <https://doi.org/10.1007/BF01186081>.
- LOW, J. W.; MWANGA, R. O. M.; ANDRADE, M.; CAREY, E.; BALL, A.-M. Tackling vitamin A deficiency with biofortified sweetpotato in sub-Saharan Africa. **Global Food Security**, v. 14, p. 23–30, set. 2017. <https://doi.org/10.1016/j.gfs.2017.01.004>.
- MCKIBBIN, R. S.; MUTTUCUMARU, N.; PAUL, M. J.; POWERS, S. J.; BURRELL, M. M.; COATES, S.; PURCELL, P. C.; TIESSEN, A.; GEIGENBERGER, P.; HALFORD, N. G. Production of high-starch, low-glucose potatoes through over-expression of the metabolic regulator *SnRK1*. **Plant Biotechnology Journal**, v. 4, n. 4, p. 409–418, 2006. <https://doi.org/10.1111/j.1467-7652.2006.00190.x>.
- MEADE, F.; BYRNE, S.; GRIFFIN, D.; KENNEDY, C.; MESITI, F.; MILBOURNE, D. Rapid Development of KASP Markers for Disease Resistance Genes Using Pooled Whole-Genome Resequencing. **Potato Research**, v. 63, n. 1, p. 57–73, mar. 2020. <https://doi.org/10.1007/s11540-019-09428-x>.
- MOISE, A. R.; AL-BABILI, S.; WURTZEL, E. T. Mechanistic Aspects of Carotenoid Biosynthesis. **Chemical Reviews**, v. 114, n. 1, p. 164–193, 8 jan. 2014. <https://doi.org/10.1021/cr400106y>.
- MOLLINARI, M.; OLUKOLU, B. A.; PEREIRA, G. da S.; KHAN, A.; GEMENET, D.; YENCHO, G. C.; ZENG, Z.-B. Unraveling the hexaploid sweetpotato inheritance using ultra-dense multilocus mapping. **G3: Genes, Genomes, Genetics**, v. 10, n. 1, p. 281–292, jan. 2020. <https://doi.org/10.1534/g3.119.400620>.
- MOLLINARI, M.; SERANG, O. Quantitative SNP genotyping of polyploids with MassARRAY and other platforms. In: BATLEY, J. (org.). **Plant Genotyping: Methods and Protocols**. Methods in Molecular Biology. New York: Springer, 2015. p. 215–241. DOI 10.1007/978-1-4939-1966-6\_17. Disponível em: [https://doi.org/10.1007/978-1-4939-1966-6\\_17](https://doi.org/10.1007/978-1-4939-1966-6_17). Acesso em: 16 nov. 2022.
- MOTSA, N. M.; MODI, A. T.; MABHAUDHI, T. Sweet potato (*Ipomoea batatas* L.) as a drought tolerant and food security crop. **South African Journal of Science**, v. 111, n. 11–12, p. 1–8, dez. 2015. <https://doi.org/10.17159/sajs.2015/20140252>.
- MUGISA, I.; KARUNGI, J.; MUSANA, P.; ODAMA, R.; ALAJO, A.; CHELANGAT, D. M.; ANYANGA, M. O.; OLOKA, B. M.; GONÇALVES DOS SANTOS, I.; TALWANA, H.; OCHWO-SSEMAKULA, M.; EDEMA, R.; GIBSON, P.; SSALI, R.; CAMPOS, H.; OLUKOLU, B. A.; DA SILVA PEREIRA, G.; YENCHO, C.; YADA, B. Combining ability and heritability analysis of sweetpotato weevil resistance, root yield, and dry matter content in sweetpotato. **Frontiers in Plant Science**, v. 13, 7 set. 2022. DOI 10.3389/fpls.2022.956936.

Disponível em: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.956936/full>. Acesso em: 15 jun. 2024.

MUKHOPADHYAY, S. K.; CHATTOPADHYAY, A.; CHAKRABORTY, I.; BHATTACHARYA, I. Crops that feed the world 5. Sweetpotato. Sweetpotatoes for income and food security. **Food Security**, v. 3, n. 3, p. 283–305, 1 set. 2011. <https://doi.org/10.1007/s12571-011-0134-3>.

MWANGA, R. O. M.; SWANCKAERT, J.; DA SILVA PEREIRA, G.; ANDRADE, M. I.; MAKUNDE, G.; GRÜNEBERG, W. J.; KREUZE, J.; DAVID, M.; DE BOECK, B.; CAREY, E.; SSALI, R. T.; UTOBLO, O.; GEMENET, D.; ANYANGA, M. O.; YADA, B.; CHELANGAT, D. M.; OLOKA, B.; MTUNDA, K.; CHIONA, M.; ... LOW, J. W. Breeding Progress for Vitamin A, Iron and Zinc Biofortification, Drought Tolerance, and Sweetpotato Virus Disease Resistance in Sweetpotato. **Frontiers in Sustainable Food Systems**, v. 5, 15 mar. 2021. DOI 10.3389/fsufs.2021.616674. Disponível em: <https://www.frontiersin.org/articles/10.3389/fsufs.2021.616674>. Acesso em: 19 jun. 2024.

NAKAMURA, K.; OHTO, M.; YOSHIDA, N.; NAKAMURA, K. Sucrose-Induced Accumulation of  $\beta$ -Amylase Occurs Concomitant with the Accumulation of Starch and Sporamin in Leaf-Petiole Cuttings of Sweet Potato 1. **Plant Physiology**, v. 96, n. 3, p. 902–909, 1 jul. 1991. <https://doi.org/10.1104/pp.96.3.902>.

NAKITTO, M.; SSALI, R. T.; JOHANNINGSMEIER, S. D.; MOYO, M.; DE KOCK, H.; BERGET, I.; OKELLO, J. J.; MAYANJA, S.; TINYIRO, S. E.; MENDES, T.; BENARD, Y.; CHELENGAT, D.; OSARU, F.; BUGAUD, C. Decision tree scoring system to guide selection for consumer preference in sweetpotato breeding trials. **Journal of the Science of Food and Agriculture**, v. 104, n. 8, p. 4615–4625, 2024. <https://doi.org/10.1002/jsfa.12883>.

OLOKA, B. M.; DA SILVA PEREIRA, G.; AMANKWAAH, V. A.; MOLLINARI, M.; PECOTA, K. V.; YADA, B.; OLUKOLU, B. A.; ZENG, Z.-B.; CRAIG YENCHO, G. Discovery of a major QTL for root-knot nematode (*Meloidogyne incognita*) resistance in cultivated sweetpotato (*Ipomoea batatas*). **Theoretical and applied genetics**, v. 134, n. 7, p. 1945–1955, jul. 2021. <https://doi.org/10.1007/s00122-021-03797-z>.

OSORIO, C. E. The Role of Orange Gene in Carotenoid Accumulation: Manipulating Chromoplasts Toward a Colored Future. **Frontiers in Plant Science**, v. 10, 4 out. 2019. DOI 10.3389/fpls.2019.01235. Disponível em: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2019.01235/full>. Acesso em: 18 jun. 2024.

PEREIRA, G. S.; AMANKWAAH, V. A.; KETAVI, M.; OLOKA, B. M.; NAIR, A. G. H.; DA MATA, A. P.; DA SILVA, C. C.; DOS SANTOS, I. G.; ROSA, J. R. B. F.; CAMPOS, H. Sweetpotato: Nutritional Constituents and Genetic Composition. In: KOLE, C. (org.). **Compendium of Crop Genome Designing for Nutraceuticals**. Singapore: Springer Nature Singapore, 2023. p. 1545–1587. DOI 10.1007/978-981-19-4169-6\_58. Disponível em: [https://link.springer.com/10.1007/978-981-19-4169-6\\_58](https://link.springer.com/10.1007/978-981-19-4169-6_58). Acesso em: 19 jun. 2024.

SCHWARZ, G. Estimating the Dimension of a Model. **The Annals of Statistics**, v. 6, n. 2, p. 461–464, 1978. .

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, jan. 2014. <https://doi.org/10.1007/s11032-013-9917-x>.

SMITH, A. M.; ZEEMAN, S. C.; SMITH, S. M. Starch degradation. **Annual Review of Plant Biology**, v. 56, p. 73–98, 2 jun. 2005. <https://doi.org/10.1146/annurev.arplant.56.032604.144257>.

SRISUWAN, S.; SIHACHAKR, D.; SILJAK-YAKOVLEV, S. The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. **Plant Science**, v. 171, n. 3, p. 424–433, 1 set. 2006. <https://doi.org/10.1016/j.plantsci.2006.05.007>.

UTOBLO, O.; AKANSAKE, D.; DORGBETOR, I.; SAAKA, J.; SWANCKAERT, J.; CAREY, E.; GEMENET, D.; CAREY, D. E. Phenotypic and QTL Analysis of the BT Population in Ghana. *In*: 17TH SWEETPOTATO BREEDERS' MEETING, 7 jun. 2018. [*S. l.: s. n.*], 7 jun. 2018.

VAN GEEST, G.; VOORRIPS, R. E.; ESSELINK, D.; POST, A.; VISSER, R. G.; ARENS, P. Conclusive evidence for hexasomic inheritance in chrysanthemum based on analysis of a 183 k SNP array. **BMC Genomics**, v. 18, n. 1, p. 585, 7 ago. 2017. <https://doi.org/10.1186/s12864-017-4003-0>.

VARGAS, P. F.; FERNANDES, A. M.; DANGO, F. J. U.; ANDRADE, M. I.; VENDRAME, L. P. C.; PEREIRA, G. S. Batata-doce: produção, consumo e formas de propagação. **Informe Agropecuário**, Hortaliças Panc: raízes e rizomas. v. 43, n. 318, p. 40–50, 2022. .

VOORRIPS, R. E.; GORT, G.; VOSMAN, B. Genotype calling in tetraploid species from bi-allelic marker data using mixture models. **BMC Bioinformatics**, v. 12, n. 1, p. 172, dez. 2011. <https://doi.org/10.1186/1471-2105-12-172>.

WANG, W.; CELTON, J.-M.; BUCK-SORLIN, G.; BALZERGUE, S.; BUCHER, E.; LAURENS, F. Skin Color in Apple Fruit (*Malus × domestica*): Genetic and Epigenetic Insights. **Epigenomes**, v. 4, n. 3, p. 13, set. 2020. <https://doi.org/10.3390/epigenomes4030013>.

WOOLFE, J. A. **Sweet potato: an untapped food resource**. 1. ed. [*S. l.*]: Cambridge University Press, 1992.

WU, S.; LAU, K. H.; CAO, Q.; HAMILTON, J. P.; SUN, H.; ZHOU, C.; ESERMAN, L.; GEMENET, D. C.; OLUKOLU, B. A.; WANG, H.; CRISOVAN, E.; GODDEN, G. T.; JIAO, C.; WANG, X.; KITAVI, M.; MANRIQUE-CARPINTERO, N.; VAILLANCOURT, B.; WIEGERT-RININGER, K.; YANG, X.; ... FEI, Z. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. **Nature Communications**, v. 9, n. 1, p. 4580, 2 nov. 2018. <https://doi.org/10.1038/s41467-018-06983-8>.

YADA, B.; BROWN-GUEDIRA, G.; ALAJO, A.; SSEMAKULA, G. N.; OWUSU-MENSAH, E.; CAREY, E.; O.M. MWANGA, R.; YENCHO, G. C. Genetic analysis and association of simple sequence repeat markers with storage root yield, dry matter, starch and  $\beta$ -carotene content in sweetpotato. **Breeding Science**, v. 67, n. 2, p. 140–150, 2017. <https://doi.org/10.1270/jsbbs.16089>.

YOON, H.; NO, J.; KIM, W.; SHIN, M. Textural character of sweet potato root of Korean cultivars in relation to chemical constituents and their properties. **Food Science and Biotechnology**, v. 27, n. 6, p. 1627–1637, 1 dez. 2018. <https://doi.org/10.1007/s10068-018-0429-7>.

ZHANG, K.; WU, Z.; WU, X.; HAN, H.; JU, X.; FAN, Y.; YANG, C.; TANG, D.; CAO, Q.; WANG, J.; LV, C. Regulatory and functional divergence among members of *Ibβfruct2*, a sweet potato vacuolar invertase gene controlling starch and glucose content. **Frontiers in Plant Science**, v. 14, 27 jun. 2023.