

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA**

LUINA RIBEIRO NOIA

**ESTUDOS AGRONÔMICOS E GENÔMICOS DE *Manihot esculenta*
Crantz: DIVERSIDADE, CORRELAÇÃO ENTRE CARACTERES E
NÚMERO MÍNIMO DE SNPs.**

**ALEGRE
2023**

LUINA RIBEIRO NOIA

**ESTUDOS AGRONÔMICOS E GENÔMICOS DE *Manihot esculenta*
Crantz: DIVERSIDADE, CORRELAÇÃO ENTRE CARACTERES E
NÚMERO MÍNIMO DE SNPs.**

Tese apresentada ao programa de Pós-Graduação em Agronomia do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Doutora em Agronomia.

Orientador: Prof. Dr. Adésio Ferreira

ALEGRE
2023

Ficha catalográfica disponibilizada pelo Sistema Integrado de
Bibliotecas - SIBI/UFES e elaborada pelo autor

N782e Noia, Luina Ribeiro, 1991-
Estudos agronômicos e genômicos de *Manihot esculenta*
Crantz: diversidade, correlação entre caracteres e número mínimo
de SNPs / Luina Ribeiro Noia. - 2023.
84 f. : il.

Orientador: Adésio Ferreira.
Coorientadora: Marcia Flores da Silva Ferreira.
Tese (Doutorado em Agronomia) - Universidade Federal do
Espírito Santo, Centro de Ciências Agrárias e Engenharias.

1. Mandioca. 2. Plantas - Melhoramento genético. I.
Ferreira, Adésio. II. Ferreira, Marcia Flores da Silva. III.
Universidade Federal do Espírito Santo. Centro de Ciências
Agrárias e Engenharias. IV. Título.

CDU: 63

LUINA RIBEIRO NOIA

**ESTUDOS AGRONÔMICOS E GENÔMICOS DE *Manihot esculenta* Crantz:
DIVERSIDADE, CORRELAÇÃO ENTRE CARACTERES E NÚMERO MÍNIMO DE
SNPS.**

Tese apresentada ao programa de Pós-Graduação em Agronomia do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Doutora em Agronomia.

Aprovada em 23 de março de 2023.

COMISSÃO EXAMINADORA:

Prof. Dr. Adésio Ferreira
Universidade Federal do Espírito Santo - Orientador

Prof^a. Dr^a. Marcia Flores da Silva Ferreira
Universidade Federal do Espírito Santo - Coorientadora

Prof. Dr. José Carlos Lopes
Universidade Federal do Espírito Santo - Examinador Interno

Prof^a. Dr^a. Camila Ferreira Azevedo
Universidade Federal de Viçosa - Examinadora Externa

Dr^a. Sheila Cristina Prucoli Posse
Incapar - Examinadora Externa

Dr^a. Carolina de Oliveira Bernardes
Universidade Federal do Espírito Santo - Examinadora Externa



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

PROTOCOLO DE ASSINATURA



O documento acima foi assinado digitalmente com senha eletrônica através do Protocolo Web, conforme Portaria UFES nº 1.269 de 30/08/2018, por
ADESIO FERREIRA - SIAPE 1573758
Departamento de Agronomia - DA/CCAE
Em 05/04/2023 às 09:49

Para verificar as assinaturas e visualizar o documento original acesse o link:
<https://api.lepisma.ufes.br/arquivos-assinados/684960?tipoArquivo=O>



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

PROTOCOLO DE ASSINATURA



O documento acima foi assinado digitalmente com senha eletrônica através do Protocolo Web, conforme Portaria UFES nº 1.269 de 30/08/2018, por
MARCIA FLORES DA SILVA FERREIRA - SIAPE 1615391
Departamento de Agronomia - DA/CCAE
Em 05/04/2023 às 10:50

Para verificar as assinaturas e visualizar o documento original acesse o link:
<https://api.lepisma.ufes.br/arquivos-assinados/685054?tipoArquivo=O>



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

PROTOCOLO DE ASSINATURA



O documento acima foi assinado digitalmente com senha eletrônica através do Protocolo Web, conforme Portaria UFES nº 1.269 de 30/08/2018, por
JOSE CARLOS LOPES - SIAPE 295710
Departamento de Agronomia - DA/CCAE
Em 05/04/2023 às 18:16

Para verificar as assinaturas e visualizar o documento original acesse o link:
<https://api.lepisma.ufes.br/arquivos-assinados/685654?tipoArquivo=O>

*Ao meu filho Vicente,
a quem sempre dedicarei
cada momento vivido.*

AGRADECIMENTOS

À Universidade Federal do Espírito Santo e ao Programa de Pós-Graduação em Agronomia pela oportunidade de realizar o doutorado.

Ao meu orientador, Prof. Adésio, por ser meu grande incentivador, por ter acreditado em mim para a realização deste trabalho, pela compreensão, empolgação e amizade.

À minha querida coorientadora, Prof^a. Marcia, por ter me incentivado nos momentos de desânimo, pelos ensinamentos, disponibilidade e carinho.

Aos membros da banca de defesa pela disponibilidade e valiosa contribuição.

Aos órgãos de fomento, CAPES, FAPES e CNPq.

Ao Incaper e Embrapa, pelas contribuições para realização da pesquisa.

A todos que contribuíram para realização deste trabalho, em especial à Natália Zardo.

A todos dos Laboratórios de Genética e Melhoramento e de Biometria da Ufes pela boa convivência e compartilhamento de experiências.

À Iasmine, por cuidar tão bem dos “nossos cafés”.

A todos os professores que contribuíram para minha formação e aos colegas de disciplinas, com os quais pude aprender muito.

Aos meus amigos da Secretaria de Pós-Graduação, Cintia, Roziane e Toninho pela maravilhosa convivência e ajuda durante a finalização e defesa da tese.

Aos meus pais, Luiz e Rosa, e meu irmão, Carlos Hartur, por todo apoio, ajuda e por tornarem minhas lutas, suas lutas.

Ao meu marido João Paulo, pelo amor, amizade, por ser meu parceiro e caminhar sempre ao meu lado.

Ao meu filho, Vicente, por me acalmar e alegrar, por ser minha força e dar sentido a tudo que faço.

A Deus, pela natureza e por conceder-me vida e saúde para admirá-la e surpreender-me todos os dias com sua perfeição.

“A natureza não faz nada em vão.”

Aristóteles

RESUMO GERAL

NOIA, Luina Ribeiro. **Estudos agronômicos e genômicos de *Manihot esculenta* Crantz: diversidade, correlação entre caracteres e número mínimo de SNPs.** 2023. Tese (Doutorado em Agronomia) - Universidade Federal do Espírito Santo, Alegre, ES, 2023.

A mandioca é uma cultura de segurança alimentar com ampla diversidade genética disponível para exploração no melhoramento da cultura. Neste trabalho, foi analisada a diversidade genética de uma população de mandioca constituída de 87 genótipos cultivados por pequenos produtores no estado do Espírito Santo usando caracteres quantitativos, qualitativos e SNPs. Também foi verificada a correlação entre os caracteres e marcadores e o número mínimo de SNPs necessário para as análises. Com os SNPs foram calculados os parâmetros genéticos heterozigosidade observada (H_o) e esperada (H_e), conteúdo de informação polimórfica (PIC) e índice de fixação (F_i), cujas médias foram de 0,46; 0,35; 0,28 e -0,34. Os valores de heterozigosidade e PIC são considerados altos diante da natureza bialélica dos marcadores e dos valores relatados na literatura. O valor de H_o maior que H_e e o F_i negativo indicam alto número de heterozigotos. Os caracteres quantitativos foram avaliados através da metodologia de modelos mistos (REML/BLUP) e as herdabilidades estimadas foram de 0,79; 0,55; 0,62 e 0,63 para Altura da Parte Aérea (HPA), Número de Raízes Totais (NRT), Peso de Raízes Totais (PESORT) e Comerciais (PESORC), respectivamente. Foram também avaliados 14 caracteres qualitativos, incluindo descritores morfológicos das raízes e parte aérea, tempo de cozimento e presença de doença, dos quais os mais variáveis foram a forma do lóbulo central, cor do pecíolo e cor externa do caule. Os genótipos com maiores PESORT apresentaram cor da polpa branca e cozimento entre zero e 20 min, sendo de grande interesse para o melhoramento da cultura. Os dendrogramas com base nos marcadores SNPs e caracteres quantitativos e qualitativos formaram 17; 9 e 10 grupos e apresentaram constituições diferentes, com exceção de alguns genótipos que mostraram similaridade nos agrupamentos por dados quantitativos e de SNPs. A correlação foi significativa entre quinze pares de caracteres, destacando-se a correlação entre os SNPs com NRT e PESORT. Foram então selecionados através de técnicas de reamostragem, conjuntos com 4; 15 e 50 SNPs cujas matrizes de distância tiveram correlação entre 0,4 e 0,56 com NRT e PESORT. Os conjuntos com quatro SNPs permitiram diferenciar os genótipos de acordo com os valores de NRT e PESORT. A análise de número mínimo de SNPs demonstrou que poucos SNPs são necessários para o cálculo de parâmetros genéticos, sendo possível recuperar os valores com apenas 50 SNPs se considerado um intervalo de confiança de 95%. Para a obtenção de

dendrogramas com os mesmos grupos formados com todos os SNPs, é necessário um número maior de marcadores, de 3000 para a população de mandioca. Este valor pode ser calculado por técnicas de reamostragem e correlação de matrizes de distância, considerando-se um valor de correlação de 0,99. As análises caracterizaram ferramentas importantes para o melhoramento da mandioca, dentre elas, uma população com diversidade molecular, morfológica e produtiva suficiente para a seleção de genótipos com características agronômicas de interesse; SNPs com correlação com as principais características de produtividade que poderão ser estudados para seleção assistida por marcadores; e o número mínimo de SNPs para avaliação da diversidade genética na população.

Palavras-chave: mandioca; melhoramento de plantas; caracteres quantitativos; descritores morfológicos; DArTseq.

GENERAL ABSTRACT

NOIA, Luina Ribeiro. **Agronomic and genomic studies of *Manihot esculenta* Crantz: diversity, character correlation and minimum number of SNPs.** 2023. Thesis (Doctorate in Agronomy) - Universidade Federal do Espírito Santo, Alegre, ES, 2023.

Cassava is a food security crop with wide genetic diversity available for research in plant breeding. In this work, the genetic diversity of a cassava population consisting of 87 genotypes cultivated by small producers in the State of Espírito Santo was evaluated using SNPs, quantitative and qualitative traits. The correlation between traits and markers and the minimum number of SNPs required for the analyzes was also verified. The genetic parameters observed (H_o) and expected (H_e) heterozygosity, polymorphic information content (PIC) and fixation index (F_i) were calculated with the SNPs, which means were 0.46; 0.35; 0.28; and -0.34, respectively. Heterozygosity and PIC values can be considered high given the biallelic nature of the markers and the values reported by other authors. H_o values greater than H_e and negative F_i indicate a high number of heterozygotes. The quantitative traits were evaluated using the mixed model methodology (REML/BLUP) and heritabilities of 0.79, 0.55, 0.62 and 0.63 were observed for Height of the Aerial Part (HPA), Number of Total Roots (NRT), Weight of Total Roots (PESORT) and Commercial Roots (PESORC), respectively. Fourteen qualitative traits were also evaluated, including morphological descriptors of roots and aerial part, cooking time and presence of disease, of which the most variable were central lobe shape, petiole color and external stem color. The genotypes with the highest PESORT had white flesh and cooking time between zero and 20 min. The dendrograms based on SNPs, quantitative and qualitative traits formed 17, 9 and 10 groups and showed different composition, with the exception of some genotypes that showed similarity in the clusters regarding quantitative and SNPs data. The correlation was significant between fifteen pairs of characters, highlighting the correlation between SNPs with NRT and PESORT. Sets with 4, 15 and 50 SNPs were selected by resampling techniques, and distance matrices computed with these sets presented correlation between 0.4 and 0.56 with NRT and PESORT. The sets with four SNPs allowed to differentiate the genotypes according to the NRT and PESORT values. The analysis of the minimum number of SNPs demonstrated that few SNPs are necessary to calculate of genetic parameters, and it is possible to recover the values with only 50 SNPs if a confidence interval of 95% is considered. To obtain dendrograms with the same groups formed with all SNPs, a greater number of markers is required, 3000 for the cassava population. This value can be calculated by resampling and distance matrix correlation

techniques, considering a correlation value of 0.99. The analyzes characterized important tools for cassava breeding, among them, a population with sufficient molecular, morphological and productive diversity for the selection of genotypes with agronomic characteristics of interest; SNPs with correlation with the main productivity traits that can be studied for marker-assisted selection; and the minimum number of SNPs to evaluate genetic diversity in the population.

Keywords: cassava; plant breeding; quantitative characters; morphological descriptors; DArTseq.

LISTA DE FIGURAS

CAPÍTULO I - Diversidade genética de *Manihot esculenta* e correlação entre caracteres quantitativos, qualitativos e SNPs

- Figura 1.** *Workflow* das análises de diversidade genética de 87 genótipos de mandioca com caracteres quantitativos e qualitativos e marcadores SNPs.25
- Figura 2.** *Workflow* das análises de correlação de 87 genótipos de mandioca com caracteres quantitativos e qualitativos e marcadores SNPs.29
- Figura 3.** *Workflow* para identificação de amostras de SNPs com alta correlação com características quantitativas e qualitativas.31
- Figura 4.** Distribuição dos valores de distância genética entre 87 genótipos de mandioca com base em marcadores SNPs (A) e caracteres quantitativos (B) e qualitativos (C).33
- Figura 5.** Dendrograma UPGMA de 87 genótipos de mandioca com base em marcadores SNPs (A) e caracteres quantitativos (B) e qualitativos (C).34
- Figura 6.** Correlação e Dispersão entre os caracteres quantitativos Altura de Plantas (HPA), Número de Raízes Totais (NRT), Peso de Raízes Comerciais (PESORC), e Peso de Raízes Totais (PESORT) medidos em 87 genótipos de mandioca.39
- Figura 7.** Gráfico de barras empilhadas de pares de caracteres qualitativos com associação significativa (p -valor $< 0,01$) pelo teste de Fisher. CorER = Cor Externa da Raiz; TextEp = Textura da Epiderme da Raiz; CorCR = Cor do Córtex da Raiz; CorP = Cor da Polpa da Raiz; Pub = Pubescência da raiz; CorC = Cor do Córtex do Caule. Os caracteres foram avaliados em 87 genótipos de mandioca.41
- Figura 8.** Boxplots dos valores das características quantitativas Número de Raízes Totais, Peso de Raízes Totais e Peso de Raízes Comerciais por classe das qualitativas Tempo de Cozimento ou Cor da Folha Apical observadas em 87 genótipos de mandioca. Os entalhes não sobrepostos dão cerca de 95% de confiança de que duas medianas diferem.42
- Figura 9.** Gráfico de dispersão do número de SNPs *versus* os valores de correlação dos SNPs mais frequentes dentre 1000 amostras de 2000 SNPs cujas matrizes de distância tiveram as maiores correlações com a matriz de distância do Número de Raízes Totais de 87 genótipos de mandioca.43
- Figura 10.** (A) Heatmap de quatro SNPs cujas matrizes de distância tiveram correlação de 0,45 com a matriz de distância do Número de Raízes Totais (NRT) de 87 genótipos de mandioca. Legenda do Heatmap: ■ homozigotos para o alelo de referência; ■ homozigotos para o alelo alternativo; ■ heterozigoto. Legenda dos genótipos e valores de NRT: ● <6,65; ● 6,65-7,00; ● 7,00-7,35, ● >7,35. (B) Boxplot do NRT por grupo de combinação de SNPs. A = alelo de referência; a = alelo alternativo.45
- Figura 11.** Boxplot dos valores de Peso de Raízes Totais (PESORT) entre os genótipos agrupados de acordo com os SNPs S2_3333213 e S2_4225917 em 87 genótipos de mandioca. RR = homozigotos para o alelo alternativo nos dois SNPs; RD = um homozigoto para o alelo alternativo e um homozigoto para o alelo de referência; HH = heterozigotos para os dois SNPs; RH = um homozigoto para o alelo alternativo e um heterozigoto; RH = um homozigoto para o alelo de referência e um heterozigoto. Os entalhes não sobrepostos dão cerca de 95% de confiança de que duas medianas diferem.46

CAPÍTULO II - Número Mínimo de SNPs para estudos de diversidade genética em *Manihot esculenta* Crantz

Figura 1. Boxplots dos valores dos parâmetros genéticos Heterozigidade Observada (H_o) e Esperada (H_e) e Índice de Fixação (F_i) calculados com diferentes números de SNPs em uma população de 87 genótipos de mandioca. 66

Figura 2. Dendrograma UPGMA da população de mandioca com base em 17816 marcadores SNPs genotipados em 87 genótipos de mandioca. Ponto de corte em 0,65. 68

Figura 3. Gráficos de curva revelando a tendência de correlação da matriz de distância com base em diferentes números de SNPs com a matriz gerada com todos os SNPs. (A) População de mandioca; (B) População simulada 1; (C) População simulada 2. As três populações são constituídas de 87 genótipos. 70

LISTA DE TABELAS

CAPÍTULO I - Diversidade genética de *Manihot esculenta* Crantz e correlação entre caracteres quantitativos, qualitativos e SNPs

Tabela 1. Região, altitude e clima dos municípios onde foram implantados experimentos com 87 genótipos de mandioca no Estado do Espírito Santo.....23

Tabela 2. Diferentes modelos para a estrutura de covariância da matriz G.28

Tabela 3. Número e frequência de genótipos por classe de 14 caracteres qualitativos avaliados em 87 genótipos de mandioca.37

CAPÍTULO II - Número Mínimo de SNPs para estudos de diversidade genética em *Manihot esculenta* Crantz

Tabela 1. Média dos parâmetros genéticos, média da distância euclidiana e número de grupos formados no agrupamento UPGMA de 87 genótipos com base em marcadores SNPs genotipados em uma população de mandioca e simulados para duas populações (PS1 e PS2).65

SUMÁRIO

1 INTRODUÇÃO GERAL	15
2 REFERÊNCIAS	18

CAPÍTULO I - Diversidade genética de <i>Manihot esculenta</i> Crantz e correlação entre caracteres quantitativos, qualitativos e SNPs	20
--	-----------

1 INTRODUÇÃO	21
2. MATERIAL E MÉTODOS	23
2.1 MATERIAL GENÉTICO.....	23
2.2 DELINEAMENTO EXPERIMENTAL	23
2.3 AVALIAÇÃO DE CARACTERES QUANTITATIVOS E QUALITATIVOS.....	24
2.4 GENOTIPAGEM VIA TECNOLOGIA DArTseq.....	24
2.5 ANÁLISES ESTATÍSTICAS	25
2.5.1 Análises de Diversidade Genética	25
2.5.2 Correlação entre caracteres.....	29
3 RESULTADOS	32
3.1 DIVERSIDADE GENÉTICA POR MARCADORES MOLECULARES	32
3.2 DIVERSIDADE GENÉTICA POR CARACTERES QUANTITATIVOS	35
3.3 DIVERSIDADE GENÉTICA POR CARACTERES QUALITATIVOS	36
3.4 CORRELAÇÃO ENTRE CARACTERES QUANTITATIVOS.....	38
3.5 ASSOCIAÇÃO ENTRE CARACTERES QUALITATIVOS	39
3.6 CORRELAÇÃO ENTRE MATRIZES DE DISTÂNCIA DOS CARACTERES QUANTITATIVOS E QUALITATIVOS E SNPS.....	42
4 DISCUSSÃO	47
4.1 DIVERSIDADE GENÉTICA	47
4.2 CORRELAÇÃO ENTRE CARACTERES	51
5 CONCLUSÕES.....	54
6 REFERÊNCIAS	55

CAPÍTULO II - Número mínimo de SNPs para estudos de diversidade genética em <i>Manihot esculenta</i> Crantz	60
--	-----------

1 INTRODUÇÃO	61
2 METODOLOGIA.....	62
2.1 CONJUNTOS DE DADOS SNPs.....	62
2.1.1 SNPs genotipados em Mandioca	62
2.1.2 SNPs simulados	63

2.2 NÚMERO MÍNIMO DE SNPs PARA O CÁLCULO DE PARÂMETROS GENÉTICOS	63
2.3 NÚMERO MÍNIMO DE SNPs PARA ANÁLISES DE AGRUPAMENTOS.....	63
2.4 NÚMERO MÍNIMO DE SNPs POR CORRELAÇÃO ENTRE MATRIZES	64
2.5 IDENTIFICAÇÃO DE SNPS ESPECÍFICOS PARA OS ESTUDOS DE DIVERSIDADE GENÉTICA	64
3 RESULTADOS.....	65
3.1 CARACTERIZAÇÃO DAS POPULAÇÕES.....	65
3.2 NÚMERO MÍNIMO DE SNPs PARA O CÁLCULO DE PARÂMETROS GENÉTICOS.....	65
3.3 NÚMERO MÍNIMO DE SNPs PARA AGRUPAMENTOS	67
3.4 NÚMERO MÍNIMO DE SNPs POR CORRELAÇÃO ENTRE MATRIZES	69
4 DISCUSSÃO	71
5 CONCLUSÕES.....	75
6 REFERÊNCIAS	76

1 INTRODUÇÃO GERAL

A mandioca (*Manihot esculenta* Crantz) é a terceira fonte mais importante de calorias nos trópicos (FAO, 2020), servindo como alimento básico para populações localizadas em muitos países, como Nigéria, Tailândia, Brasil e Indonésia, que são os maiores produtores mundiais (VISSÉS; SENTELHA; PEREIRA, 2018; FAOSTAT, 2023). Sua importância na segurança alimentar deve-se às suas características benéficas, como as raízes nutritivas, capacidade de se adaptar a quase todos os tipos de clima, tolerância à seca e a solos inférteis, onde muitas outras culturas apresentam desempenho ruim, capacidade de se recuperar de doenças e ataques de pragas, além de não requerer cuidados exigentes (MORANTE et al., 2010; BYJU; SUJA, 2020; DE OLIVEIRA et al., 2022). Além do importante papel que a mandioca desempenha na segurança alimentar, há uma demanda crescente por raízes de mandioca pelas indústrias de amido, alimentos, ração animal e etanol (CEBALLOS et al., 2007). Desta forma, a mandioca tornou-se uma cultura polivalente que responde às prioridades dos países em desenvolvimento, às tendências da economia global e ao desafio das mudanças climáticas (HOWELER; LUTALADIO; THOMAS, 2013).

O Brasil é o principal centro de diversidade da mandioca (BREDESON et al., 2016; ALVES-PEREIRA et al., 2019). Como a mandioca é produzida principalmente por pequenos produtores e a maioria dos seus consumidores vive em países em desenvolvimento, a cultura não recebeu o cultivo intenso e o melhoramento genético realizado nas culturas de cereais, como milho, trigo e arroz (BART; TAYLOR, 2017; AMELEWORK et al., 2021). Portanto, a variabilidade genética da espécie ainda é muito conservada, com potencial para o melhoramento da cultura, visando aumento produtivo, adaptabilidade e características favoráveis ao consumo (AMELEWORK; BAIRU, 2022).

A análise da diversidade genética de uma população pode ser realizada de diversas formas, sendo que o uso conjunto de diferentes tipos de caracteres e marcadores – morfológicos, agrônômicos e moleculares – permite obter uma visão abrangente da diversidade real disponível (OLIVEIRA; OLIVEIRA FILHO; SANTOS, 2015).

As características economicamente mais importantes nas espécies agrícolas são as quantitativas. As análises destas características têm sido continuamente reinventadas, de forma a atender as necessidades que surgem com os avanços do melhoramento de plantas (BERNANDO, 2020). Exemplo disso são as análises via modelos mistos (REML/BLUP). O uso da máxima verossimilhança restrita (REML) para obtenção dos componentes de variância seguido da melhor predição linear imparcial (BLUP) é o método mais eficiente para

estimativa de parâmetros genéticos e predição de valores genotípicos, principalmente em dados desbalanceados, gerando resultados com acurácia e não viesados (PIEPHO et al., 2008; RESENDE et al., 2012). A metodologia assegura uma estimativa mais precisa da diversidade genética, pois permite o uso dos valores genotípicos ao invés dos fenotípicos, sem a influência dos efeitos ambientais identificáveis (RESENDE et al., 2012; CARVALHO et al., 2020; DE SÁ et al., 2021).

Além dos caracteres quantitativos, a avaliação de caracteres qualitativos é de grande importância para o melhoramento de plantas, pois, em conjunto com os primeiros, proporciona a diferenciação e identificação de materiais, contribuindo para o intercâmbio de germoplasma e para a organização e manutenção dos recursos genéticos da espécie, considerando as atividades de coleta, caracterização e conservação (FUKUDA et al., 2003; SILVA et al., 2017). Além disso, características qualitativas podem contribuir para o processo de seleção, de forma a gerar variedades com características que são preferidas pelos produtores ou consumidores. No caso da mandioca, a cor da polpa e da casca, rugosidade e presença de pedúnculo são exemplos de características importantes para o comércio e manejo da cultura (LORENZI et al., 2012).

Por fim, os marcadores moleculares, que são aqueles detectados diretamente no DNA, possibilitam a avaliação da diversidade como uma característica genômica, independente de estar ou não correlacionada com a expressão fenotípica. Eles são estáveis e detectáveis em todos os tecidos, independentemente do crescimento, diferenciação ou desenvolvimento da planta e não são confundidos por efeitos ambientais, pleiotrópicos e epistáticos (MONDINI; NOORANI; PAGNOTTA, 2009). Dentre os marcadores moleculares, os SNPs (Polimorfismos de Nucleotídeo Único) têm sido amplamente usados em diversas culturas para detecção da diversidade genética, com a vantagem de terem alta abundância nos genomas, alto nível de reprodutibilidade e formatos de detecção de alto rendimento (MAMMADOV et al., 2012; DAR; MAHAJAN; SHARMA, 2019), o que também tem levado à detecção de regiões associadas a características agrônômicas de interesse.

Em mandioca, as análises com marcadores SNPs são ainda recentes, tendo sido iniciadas com ESTs (*expressed sequence tags*), com a detecção de poucos SNPs, entre 2.000 e 3.000 (FERGUSON et al., 2011; POOTAKHAM et al., 2014) e, posteriormente, atingindo até 56 mil SNPs de qualidade (RABBI et al., 2015) com o advento das técnicas do sequenciamento de nova geração (NGS) (CHUNG et al., 2017). Entretanto, sequenciar SNPs de alta densidade em um grande número de indivíduos acaba aumentando os custos e o tempo analítico (MORIN et al. 2004; NAZARENO et al., 2017). O desenvolvimento de painéis com

número reduzido de SNPs pode tornar as análises mais acessíveis, o que é especialmente importante para culturas como a mandioca, cujas pesquisas são realizadas principalmente por instituições públicas de países em desenvolvimento (AMELEWOR e BAIRU, 2021), com limitação de recursos. Porém, para a maioria das espécies, ainda é necessário identificar o número e as características ideais desses marcadores para os estudos de diversidade.

Neste estudo, foram caracterizadas ferramentas importantes para o melhoramento da mandioca, destacando-se a diversidade genética de uma população de mandioca constituída de genótipos cultivados por pequenos produtores no Estado do Espírito Santo usando caracteres quantitativos, qualitativos e SNPs; a correlação entre os caracteres e marcadores, incluindo a identificação de SNPs com correlação com as principais características de produtividade que poderão ser estudados para seleção assistida por marcadores; e o número mínimo de SNPs para avaliação da diversidade genética na população por diferentes técnicas.

2 REFERÊNCIAS

- ALVES-PEREIRA, A.; CLEMENT, C. R.; PICANÇO-RODRIGUES, D.; et al. A population genomics appraisal suggests independent dispersals for bitter and sweet manioc in Brazilian Amazonia. **Evolutionary applications**, v. 13, n. 2, p. 342-361, 2020. <https://doi.org/10.1111%2Feva.12873>
- AMELEWORK, A.B.; BAIRU, M.W. Advances in Genetic Analysis and Breeding of Cassava (*Manihot esculenta* Crantz): A Review. **Plants (Basel)**, v. 11, n. 1617, 2022. <https://doi.org/10.3390/plants11121617>
- AMELEWORK, A.B.; BAIRU, M.W.; MAEMA, O.; VENTER, S.L.; LAING, M. Adoption and Promotion of Resilient Crops for Climate Risk Mitigation and Import Substitution: A Case Analysis of Cassava for South African Agriculture. **Front. Sustain. Food Syst.**, v. 5, n. 617783, 2021. <https://doi.org/10.3389/fsufs.2021.617783>
- BART, R.S.; TAYLOR, N.J. New opportunities and challenges to engineer disease resistance in cassava, a staple food of African small-holder farmers. **PLoS Pathog**, v. 13, n. 5, 2017. <https://doi.org/10.1371/journal.ppat.1006287>
- BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v. 125, p. 375–385 (2020). <https://doi.org/10.1038/s41437-020-0312-1>
- BREDESON, J. V.; LYONS, J. B.; PROCHNIK, S. E.; et al. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. **Nature**, v. 34, p. 562-571, 2016. <https://doi.org/10.1038/nbt.3535>
- BYJU, G.; SUJA, G. Mineral nutrition of cassava. **Advances in Agronomy**, 159, 169-235, 2020. <https://doi.org/10.1016/bs.agron.2019.08.005>
- CARVALHO, M. S.; FERREIRA, M. F. D. S.; OLIVEIRA, W. B. D. S.; et al. Genetic diversity and population structure of *Euterpe edulis* by REML/BLUP analysis of fruit morphology and microsatellite markers. **Crop Breeding and Applied Biotechnology**, v. 20, n. 4, 2020. <https://doi.org/10.1590/1984-70332020v20n4a61>
- CEBALLOSH., FREGENEM., PÉREZJ.C., et al. Cassava genetic improvement. In: KANG, M. S.; PRIYADARSHAN, P. M. (ed.) **Breeding major food staples**. Ames, IA: Blackwell Publ., 2007. p. 365–391.
- DAR, A. A.; MAHAJAN, R.; SHARMA, S. Molecular markers for characterization and conservation of plant genetic resources. **Indian J Agric Sci**, v. 89, n. 11, 2019.
- DE OLIVEIRA, G. M.; SANTOS, J. D. O.; DOS SANTOS, C. B.; et al. Rise in temperature increases growth and yield of *Manihot* sp. plants. **Research, Society and Development**, v. 11, n. 9, 2022. <http://dx.doi.org/10.33448/rsd-v11i9.29891>
- DE SA, S. F.; DOS SANTOS, L. C. A.; DA CONCEIÇÃO, L. D. S.; et al. Genetic diversity via REML-BLUP of ex situ conserved macauba [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] ecotypes. **Genetic Resources and Crop Evolution**, v. 68, n. 8, p. 3193-3204, 2021.
- FAO (FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS). **Cassava processing: cassava wet flour**. Gana, 2020.
- FAOSTAT. (2023). **Food and Agriculture Organization of the United Nations Statistics**. Disponível em: <<http://www.fao.org/faostat/en/#data/QC>>. Acesso em: 12 jan. 2023.

- FERGUSON, M. E., HEARNE, S. J., CLOSE, T. J., et al. Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. **Theoretical and Applied Genetics**, v. 124, n. 4, p. 685–95, 2011. <https://doi.org/10.1007/s00122-011-1739-9>
- FUKUDA, W. M. G.; FUKUDA, C.; DIAS, M. C.; et al. **Cultivares de mandioca recomendadas para o Brasil**. Cruz das Almas: Embrapa, 2003. 8p.
- HOWELER, R.; LUTALADIO, N.; THOMAS, G. **Save and Grow: Cassava: A Guide to Sustainable Production Intensification**. Food and Agriculture Organization of the United Nations Rome, Italy, 2013.
- LORENZI, J. O. Mandioca. 2. ed. Campinas: CATI, 2012. 129p.
- MAMMADOV, J.; AGGARWAL, R.; BUYYARAPU, R.; KUMPATLA, S. SNP markers and their impact on plant breeding. **Int. J. Plant Genomics**, v. 2012; n.728398, 2012. <https://doi.org/10.1155/2012/728398>
- MONDINI, L., NOORANI, A., & PAGNOTTA, M. A. Assessing plant genetic diversity by molecular tools. **Diversity**, v. 1, n. 1, p. 19-35, 2009. <https://doi.org/10.3390/d1010019>
- MORANTE, N., SÁNCHEZ, T., CEBALLOS, H., et al. Tolerance to postharvest physiological deterioration in cassava roots. **Crop Science**, v. 50, n. 4, p. 1333-1338, 2010. <https://doi.org/10.2135/cropsci2009.11.0666>
- MORIN, P. A.; MARTIEN, K. K.; TAYLOR, B. L. Assessing the statistical power of SNPs for population structure and conservation studies. **Mol. Ecol. Resour.**, v. 9, n. 6, 2009. <https://doi.org/10.1111/j.1755-0998.2008.02392.x>
- NAZARENO, A. G.; BEMMELS, J. B.; DICK, C. W.; LOHMANN, L. G. Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. **Molecular Ecology Resources**, 17, n. 11, 2017. <https://doi.org/10.1111/1755-0998.12654>
- OLIVEIRA, E. J., OLIVEIRA FILHO, O. S., & SANTOS, V. S. Classification of cassava genotypes based on qualitative and quantitative data. **Genetics and Molecular Research**, v. 14, n. 1, p. 906-924, 2015. <https://doi.org/10.4238/2015.february.2.14>
- PIEPHO, H. P., RICHTER, C., & WILLIAMS, E. Nearest neighbour adjustment and linear variance models in plant breeding trials. **Biometrical Journal**, v. 50, n. 2, p. 164-189, 2008. <https://doi.org/10.1002/bimj.200710414>
- POOTAKHAM, W.; SHEARMAN, J. R.; RUANG-AREERATE, P.; et al. Large-scale SNP discovery through RNA sequencing and SNP genotyping by targeted enrichment sequencing in cassava (*Manihot esculenta* Crantz). **Plos One**, v. 9, p. 1-19, 2014. <https://doi.org/10.1371/journal.pone.0116028>
- RESENDE, M. D. V.; SILVA, F. F.; LOPES, P. S.; et al. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa: Universidade Federal de Viçosa/Departamento de Estatística. 2012. 291 p.
- SILVA, R. S.; CUNHA, E.; de FARIAS NETO, J. T.; et al. Selection of morphoagronomic descriptors for the characterization of accessions of cassava of the Eastern Brazilian Amazon. **Genetics and Molecular Research**, v. 16, n. 2, 2017. <http://dx.doi.org/10.4238/gmr16029595>
- VISSÉS, F. D. A.; SENTELHAS, P. C.; PEREIRA, A. B. Yield gap of cassava crop as a measure of food security - an example for the main Brazilian producing regions. **Food Sec.**, v. 10, 1191–1202, 2018. <https://doi.org/10.1007/s12571-018-0831-2>

CAPÍTULO I

Diversidade genética de *Manihot esculenta* Crantz e correlação entre caracteres quantitativos, qualitativos e SNPs

Resumo: A mandioca é uma cultura de segurança alimentar e seu melhoramento genético tem se tornado cada vez mais urgente diante das mudanças climáticas. Para que o melhoramento ocorra de forma bem sucedida é fundamental a identificação da diversidade genética por diferentes tipos de caracteres e marcadores, como os morfoagronômicos e moleculares, detectando a diversidade real disponível de forma abrangente. Neste trabalho, foi estudada a diversidade genética de uma população de mandioca constituída de 87 genótipos cultivados por pequenos produtores no estado do Espírito Santo usando caracteres quantitativos, qualitativos e SNPs e foi verificada a correlação entre os caracteres e marcadores. Com os SNPs foram calculados os parâmetros genéticos heterozigosidade observada (H_o) e esperada (H_e), conteúdo de informação polimórfica (PIC) e índice de fixação (F_i), cujas médias foram de 0,46; 0,35; 0,28 e -0,34, respectivamente. Os caracteres quantitativos foram analisados através da metodologia de modelos mistos (REML/BLUP) e foram observadas herdabilidades de 0,79; 0,55; 0,62 e 0,63 para Altura da Parte Aérea (HPA), Número de Raízes Totais (NRT), Peso de Raízes Totais (PESORT) e Comerciais (PESORC), respectivamente. Também foram avaliados 14 caracteres qualitativos, incluindo descritores morfológicos das raízes e parte aérea, tempo de cozimento e presença de doença, dos quais os mais variáveis foram a forma do lóbulo central, cor do pecíolo e cor externa do caule, com 8; 7 e 6 classes. Os genótipos com maiores PESORT apresentaram cor da polpa branca e cozimento entre zero e 20 min, sendo de grande interesse para o melhoramento da cultura. Os dendrogramas calculados com base nos marcadores SNPs, quantitativos e qualitativos formaram 17; 9 e 10 grupos e apresentaram constituição diferente, à exceção de alguns genótipos que mostraram similaridade nos agrupamentos por dados quantitativos e de SNPs. A correlação foi significativa entre quinze pares de caracteres, destacando-se a correlação entre os SNPs com NRT e PESORT. Dentre os SNPs, foram selecionados através de técnicas de reamostragem, conjuntos com 4, 15 e 50 SNPs cujas matrizes de distância tiveram correlação entre 0,4 e 0,56 com NRT e PESORT. Os resultados indicaram ampla diversidade genética na população avaliada, com resultados diversificados de acordo com o tipo de caractere ou marcador, o que destaca a importância de se avaliar diferentes características nas análises de diversidade.

Palavras-chave: mandioca; melhoramento de plantas; modelos mistos; DArTseq

1 INTRODUÇÃO

Diante das previsões de mudanças climáticas, olhares de todo o mundo têm se voltado para a mandioca, como cultura de segurança alimentar (PUSHPALATHA, GANGADHARAN, 2020). A planta, de raiz comestível, apresenta alta rusticidade, se desenvolvendo bem sob diversas condições, inclusive em solos pobres, áridos e em climas quentes (BYJU; SUJA, 2020; DE OLIVEIRA et al., 2022). Assim, em locais onde culturas mais exigentes como arroz, milho e trigo, não produziriam é possível cultivar a mandioca, nutrindo populações pobres de países em desenvolvimento (HOWELER; LUTALADIO; THOMAS, 2013).

A produtividade da mandioca, entretanto, ainda está abaixo do seu potencial em alguns países como o Brasil, em que são produzidas cerca de 15 t ha^{-1} , embora a cultura possa produzir até 60 t ha^{-1} (BESTER et al., 2022). Para elevar a produtividade, é necessário, além de outros fatores, o cultivo de variedades melhoradas com alto potencial produtivo e bem adaptadas.

O melhoramento da mandioca, assim como o de qualquer cultura, depende inicialmente da existência de uma população geneticamente diversificada em relação à característica a ser melhorada, mas também morfológicamente e molecularmente (SWARUP et al., 2021). A análise com diferentes tipos de caracteres e marcadores – morfológicos, agrônômicos e moleculares – permite obter uma visão abrangente da diversidade real disponível (OLIVEIRA; OLIVEIRA FILHO; SANTOS, 2015).

As características economicamente mais importantes nas espécies agrícolas são as quantitativas. As análises destas características têm sido continuamente reinventadas, de forma a atender as necessidades que surgem com os avanços do melhoramento de plantas (BERNARDO, 2020). Exemplo disso são as análises via modelos mistos (REML/BLUP). O emprego da máxima verossimilhança restrita (REML) para obtenção dos componentes de variância seguido da melhor predição linear imparcial (BLUP) é o método mais eficiente para estimativa de parâmetros genéticos e predição de valores genotípicos, principalmente em dados desbalanceados, gerando resultados com acurácia e não viesados (PIEPHO et al., 2008; RESENDE et al., 2012). A metodologia assegura uma estimativa mais precisa da diversidade genética, pois permite o uso dos valores genotípicos ao invés dos fenotípicos, sem a influência dos efeitos ambientais identificáveis (RESENDE et al., 2012; CARVALHO et al., 2020; DE SÁ et al., 2021).

Além dos caracteres quantitativos, a avaliação de caracteres qualitativos é de grande importância para o melhoramento de plantas, pois, em conjunto com os primeiros, proporciona a diferenciação e identificação de materiais, contribuindo para o intercâmbio de germoplasma e para a organização e manutenção dos recursos genéticos da espécie, considerando as atividades de coleta, caracterização e conservação (FUKUDA et al., 2003; SILVA et al., 2017). Além disso, características qualitativas podem contribuir para o processo de seleção, de forma a gerar variedades com características que são preferidas pelos produtores ou consumidores. No caso da mandioca, a cor da polpa e da casca, rugosidade e presença de pedúnculo são exemplos de características importantes para o comércio e manejo da cultura (LORENZI et al., 2012).

Por fim, os marcadores moleculares, que são aqueles detectados diretamente no DNA, possibilitam a avaliação da diversidade como uma característica genômica, independente de estar ou não correlacionada com a expressão fenotípica. Eles são estáveis e detectáveis em todos os tecidos, independentemente do crescimento, diferenciação ou desenvolvimento da planta e não são confundidos por efeitos ambientais, pleiotrópicos e epistáticos (MONDINI; NOORANI; PAGNOTTA, 2009). Dentre os marcadores moleculares, os SNPs (Polimorfismos de Nucleotídeo Único) têm sido amplamente usados em diversas culturas para detecção da diversidade genética com a vantagem de apresentarem alta abundância nos genomas, alto nível de reprodutibilidade e formatos de detecção de alto rendimento (MAMMADOV et al., 2012; DAR; MAHAJAN; SHARMA, 2019), o que também tem levado à detecção de regiões associadas a características agronômicas de interesse.

Detectada a diversidade genética, a análise de correlação e associação entre os diferentes tipos de caracteres e marcadores pode facilitar e reduzir os custos dos programas de melhoramento ao permitir a aplicação de técnicas como a seleção indireta e a seleção assistida por marcadores moleculares (DINIZ; OLIVEIRA, 2019; KUMAWAT et al., 2020).

Assim, o objetivo deste trabalho foi avaliar a diversidade genética de uma população de mandioca constituída de genótipos cultivados por pequenos produtores no estado do Espírito Santo usando caracteres quantitativos, qualitativos e SNPs e verificar a correlação entre os caracteres e marcadores.

2. MATERIAL E MÉTODOS

2.1 MATERIAL GENÉTICO

Os genótipos foram obtidos de pequenos produtores que cultivam a mandioca por várias gerações no estado do Espírito Santo, nos municípios de Castelo, Cachoeiro de Itapemirim, Jerônimo Monteiro, Alegre, Mimoso do Sul, Divino de São Lourenço, Conceição do Castelo, Venda Nova do Imigrante, Afonso Cláudio, Domingos Martins, Sooretama, São Mateus, Jaguaré, Nova Venécia e Aracruz. Foram escolhidos genótipos de mandioca de mesa que apresentavam características agronômicas de interesse (ANEXO I).

2.2 DELINEAMENTO EXPERIMENTAL

Seis experimentos foram implantados nos anos de 2018 e 2019 nos municípios de Mimoso do Sul, Venda Nova do Imigrante e Linhares no Estado do Espírito Santo, cujos climas e altitudes são descritos na Tabela 1. Cada época de plantio foi considerada um ambiente, totalizando seis ambientes.

Tabela 1. Região, altitude e clima dos municípios onde foram implantados experimentos com 87 genótipos de mandioca no estado do Espírito Santo

	Mimoso do Sul	Venda Nova	Linhares
Região	Sul	Serrana	Norte
Altitude Média	240 m	950 m	40 m
Clima	Tropical Chuvoso	Mesotérmico	Tropical Chuvoso
Inverno	Seco	Seco	Seco
Temperatura Média	24,6 °C	18,5 °C	24,3 °C
Índice Pluviométrico	1.376,4 mm	1.363,7 mm	1.278,9 mm

Fonte: Adaptado de INCAPER (2023)

O delineamento experimental utilizado foi o de blocos aumentados (DBA), sendo 84 genótipos (tratamentos não comuns) e três tratamentos comuns, distribuídos em três blocos com cinco plantas por parcela. Manivas de 15 a 20 cm foram plantadas em linhas durante a estação chuvosa nas regiões. Os espaçamentos entre linhas e plantas foram de 1,0 e 0,7 m, respectivamente. As plantas foram colhidas após 12 meses após o plantio.

2.3 AVALIAÇÃO DE CARACTERES QUANTITATIVOS E QUALITATIVOS

Os caracteres quantitativos avaliados foram: a) Altura da parte aérea (HPA) – determinada pela medição do material vegetal imediatamente após o corte a 20 cm do solo; b) Número total de raízes tuberosas (NRT); c) Peso de raízes comerciais (PESORC) – determinado pela pesagem de todas as raízes comerciais colhidas na parcela (foram consideradas raízes comerciais aquelas com comprimento e diâmetro maior ou igual a 0,20 m e 0,03 m, respectivamente); d) Peso total de raízes (PESORT) - determinado pela pesagem de todas as raízes, sem descarte;

A avaliação das características qualitativas foi expressa em categorias (Tabela 3), estabelecendo um valor numérico para cada classe. Foram analisadas a cor externa da raiz (CorER), a cor do córtex da raiz (CorCR), a cor da polpa (CorP), a textura da epiderme (TextEp), a presença de pedúnculo (PrePed), cor da folha apical (CorFA), a pubescência (Pub), a forma do lóbulo central (ForLC), a cor do pecíolo (CorPe), a cor do córtex do caule (CorC), a cor externa do caule (CorEC), a floração (Flor) e a presença de doença foliar (DoecFol), segundo descritores morfológicos e agrônômicos da cultura (FUKUDA E GUEVARA, 1998), e o tempo de cozimento (Coz), segundo metodologia proposta por Pereira et al. (1985), modificada por Borges et al., (2002).

2.4 GENOTIPAGEM VIA TECNOLOGIA DArTseq

Os tecidos foliares dos genótipos analisados foram coletados e enviados para a Intertek para a extração do DNA. A genotipagem foi realizada via metodologia DArTseqTM na *Diversity Array Technology Pty Ltd.* em Canberra-Austrália, usando uma combinação de sequenciamento de próxima geração HiSeq 2500 (Illumina, EUA) com a tecnologia DArTseqTM, conforme descrito por Sansaloni et al. (2020). As sequências genômicas foram alinhadas ao genoma de referência de mandioca v6.1. Os dados genotipados foram imputados a um painel de 388.349 SNPs usando o Beagle 5.0 algoritmo (BROWNING et al. 2018) e um painel de referência de 20.733, principalmente haplótipos de mandioca da África Oriental, derivados de uma combinação de genotipagem por sequenciamento e DArTseqTM. Os genótipos foram então codificados como 0 (homozigoto para o alelo de referência), 1 (heterozigoto) e 2 (homozigoto para o alelo alternativo). Os marcadores SNPs com mais de 10% de dados ausentes e/ou com uma frequência do alelo menos comum (MAF) abaixo de 5% foram descartados.

2.5 ANÁLISES ESTATÍSTICAS

2.5.1 Análises de Diversidade Genética

A diversidade genética da população foi avaliada com os caracteres quantitativos, qualitativos e marcadores SNPs, realizando análises de parâmetros genéticos, distância ou dissimilaridade entre os genótipos e de agrupamentos (Figura 1).

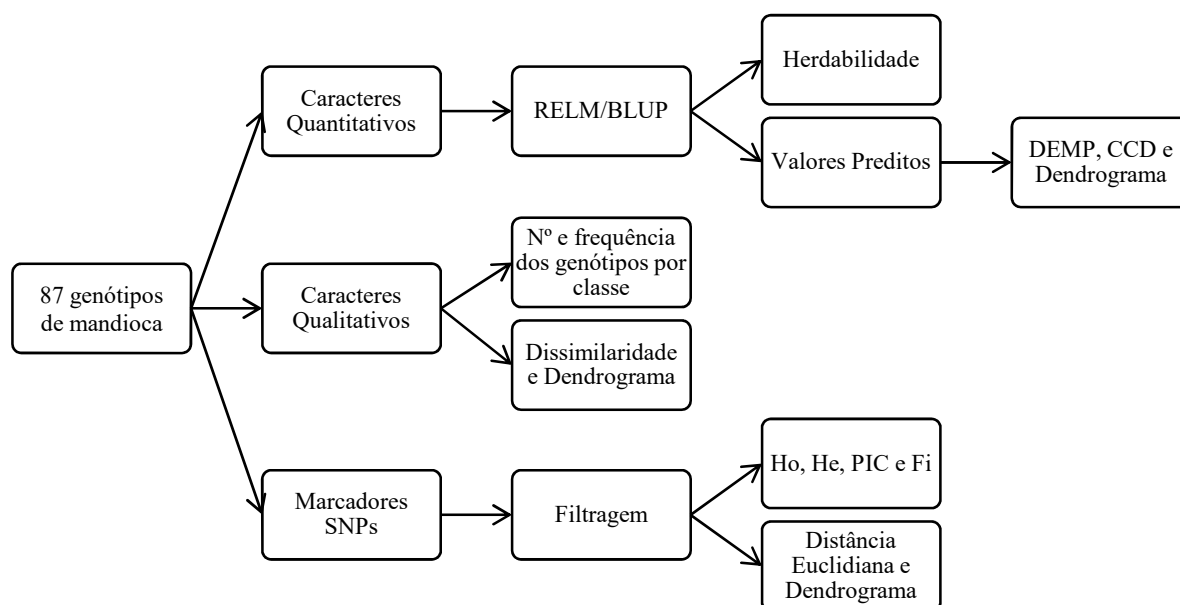


Figura 1. *Workflow* das análises de diversidade genética de 87 genótipos de mandioca com caracteres quantitativos e qualitativos e marcadores SNPs.

2.5.1.1 Marcadores SNPs

Os marcadores SNPs de qualidade resultantes após a filtragem foram armazenados em uma matriz SNPs (colunas) *versus* genótipos (linhas) e a matriz foi usada para o cálculo de parâmetros genéticos, distância euclidiana e dendrograma.

Os parâmetros genéticos analisados foram heterozigidade observada e esperada (Ho e He), conteúdo de informação polimórfica (PIC) e índice de fixação por endogamia (Fi), ambos calculados com a função “popgen” do pacote *snpReady* no software RStudio.

Para o cálculo da distância euclidiana entre os genótipos, primeiramente a matriz contendo os dados SNPs foi transformada em um objeto *genlight* usando a função “as.genlight” do pacote *adegenet*. O objeto *genlight* foi usado para o cálculo da matriz de

distância euclidiana com a função “gl.dist.ind” do pacote *dartR* no software R (R CORE TEAM, 2022). A função dimensionou as distâncias para cair no intervalo entre 0 e 1.

A distância euclidiana foi utilizada, primeiramente, por ser uma distância métrica. Distâncias não métricas são potencialmente problemáticas porque não podem necessariamente ser representadas de forma única em um espaço. Tentar representar os indivíduos em um espaço definido por coordenadas cartesianas pode causar distorções consideráveis, o que pode levar a interpretações equivocadas. A distância euclidiana permite a representação inequívoca dos indivíduos em um espaço definido pelas conhecidas coordenadas cartesianas, sem qualquer distorção, como acontece em outras distâncias métricas, ou seja, as distâncias na matriz de distância original e as distâncias no espaço reduzido coincidem na distância euclidiana. Além disso, a distância é simétrica na escolha da codificação para referência e alelo alternativo (IAE, 2022). A distância euclidiana é definida por:

$$d(A, B) = \sqrt{\sum_{i=1}^L (x_i - y_i)^2}$$

em que x_i e y_i são as contagens do alelo alternativo no locus i para os indivíduos A e B , respectivamente, e L é o número de *loci*.

A distribuição dos valores das distâncias euclidianas entre os genótipos foi apresentada na forma de um histograma plotado usando a função “hist” do software R.

O dendrograma foi calculado com base na matriz de distância euclidiana pelo método de ligação média entre grupos (UPGMA) usando a função “hclust” e plotado com a função “plot” do software R. Para a determinação do número de grupos formados foi usado o critério estatístico de Mojena (MOJENA, 1977), calculado no software Genes (CRUZ, 2013). O ponto de corte foi então inserido no dendrograma com a função “abline” do software R. O método UPGMA baseia-se na média das medidas de distância entre os genótipos, evitando valores máximos e mínimos extremos (CRUZ et al., 2011).

2.5.1.2 Caracteres Quantitativos

Os dados fenotípicos foram analisados usando a metodologia de modelos mistos (RELM/BLUP), conforme descrito por Barbiero (2022), para obtenção dos componentes de variância via REML e dos valores genéticos preditos via BLUP. Os modelos estatísticos foram baseados na estratégia de dois estágios para a análise de dados MET (*Multi-Environment Trial*). No primeiro estágio, os ensaios individuais foram analisados com

modelos incluindo termos para efeitos de delineamento e ambientes. A partir dessas análises de ensaios individuais, as médias e pesos ajustados foram inseridos no segundo estágio, onde um modelo ajusta o efeito dos genótipos por meios ambientais, usando os pesos estimados na primeira etapa.

Estágio 1: Os dados fenotípicos foram analisados individualmente dentro de cada combinação de ano e local por meio da técnica de modelos mistos. Os valores de componentes de variância foram calculados baseando-se no método de máxima verossimilhança restrita (REML) e os valores genotípicos foram estimados via o melhor estimador linear não viciado (BLUE). O modelo generalista aplicado é apresentado a seguir na equação [1]:

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{X}\mathbf{g} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \quad [1]$$

em que \mathbf{y} é o vetor ($n \times 1$) dos valores fenotípicos (dados coletados em campo); n é o número de parcelas observadas; μ é o valor da média geral; \mathbf{g} é o vetor ($i \times 1$) de efeito fixo com i genótipos; \mathbf{b} é o vetor ($j \times 1$) de efeito aleatório com k blocos, com $\mathbf{b} \sim N(\mathbf{0}, \sigma_b^2 \mathbf{I}_j)$, em que σ_b^2 é a variância de blocos; e $\boldsymbol{\varepsilon}$ é o vetor de efeito residual, com $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n)$, em que σ_ε^2 é a variância do resíduo. \mathbf{X} e \mathbf{Z} representam as matrizes de incidência para seus respectivos efeitos de genótipo e bloco; $\mathbf{1}_n$ é um vetor de uns; e \mathbf{I}_j e \mathbf{I}_n são matrizes identidade com suas respectivas ordens.

Estágio 2: Com os valores BLUEs do primeiro estágio e seus pesos, iniciou-se o desenvolvimento do modelo MET, apresentado na equação [2]:

$$\mathbf{Y} = \mu \mathbf{1}_n + \mathbf{X}\mathbf{E} + \mathbf{Z}\mathbf{ge} + \boldsymbol{\varepsilon}_1 \quad [2]$$

em que \mathbf{Y} é o vetor ($il \times 1$) de médias ajustadas do primeiro estágio, de i genótipos dentro de cada ambiente l ; μ é a média geral; \mathbf{E} é o vetor ($l \times 1$) de efeitos fixos de ambientes; \mathbf{ge} é o vetor ($il \times 1$) de efeitos genéticos aleatórios de genótipos i dentro de ambientes l , com $\mathbf{ge} \sim N(\mathbf{0}, \mathbf{G})$; e $\boldsymbol{\varepsilon}_1$ é o vetor ($il \times 1$) de efeito residual, com $\boldsymbol{\varepsilon}_1 \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n)$. \mathbf{G} é a matriz VCOV (Variância-Covariância) genética ($l \times l$) para o efeito de genótipos i dentro de ambientes l (Tabela 2), onde os elementos das diagonais são as variâncias genéticas para ambientes individuais e os elementos fora da diagonal são as covariâncias genéticas entre pares de ambientes. \mathbf{X} e \mathbf{Z} são as matrizes de incidência para seus respectivos efeitos; e $\mathbf{1}_n$ é um vetor de uns.

Tabela 2. Diferentes modelos para a estrutura de covariância da matriz G.

Modelo	va (g_{ih})	cov ($g_{ih}; g_{ih'}$)	Descrição
ID	$\sigma_g^2 + \sigma_{ge}^2$	0	Variação genética idêntica.
CORH	$\sigma_{gh}^2 + \sigma_{geh}^2$	$\rho\sigma_h\sigma_{h'}$	Modelo de correlação, forma de variância heterogênea.
FA	$\lambda_{1h}^2 + \Psi_h$	$\lambda_{1h}\lambda_{1h'}$	Modelo analítico de fator de primeira ordem.

σ_g^2 e σ_{ge}^2 : componentes de variância para efeito principal do genótipo e interação $G \times E$, respectivamente; σ_{gh}^2 e σ_{geh}^2 : variância genética específica do ambiente para efeito principal do genótipo e interação $G \times E$; λ_{1h} e $\lambda_{1h'}$: primeiros carregamentos do fator em ambientes h e h' ; ρ : correlação genética; σ_h e $\sigma_{h'}$: desvio padrão dos ambientes h e h' .

A seleção da estrutura de covariância da matriz G foi realizada utilizando os critérios: AIC - Critérios de informação de Akaike (SAKAMOTO et al., 1986) e BIC - Critério de informação Bayesiano (SCHWARZ, 1978). Foi considerado o modelo mais adequado aquele que equilibrava o ajuste dos dados e a complexidade do modelo. As equações de AIC [a] e BIC [b] foram:

$$AIC = -2 \ln L_{MAX} + 2p \text{ [a]}$$

$$BIC = -2 \ln L_{MAX} + \ln(N) p \text{ [b]}$$

em que L_{MAX} é a verossimilhança máxima; N é o número total das observações; p é o número de parâmetros da matriz de VCOC em G (PIEPHO, 2000).

Os coeficientes de herdabilidade individual de sentido amplo (h^2) foram estimados de acordo com Cullis et al., (2006), utilizando a equação: $h^2 = 1 - \left[\frac{PEV}{2 \times \hat{\sigma}_g^2} \right]$; em que PEV é a variância média do erro predito e $\hat{\sigma}_g^2$ é a estimativa do valor genético. Os coeficientes de variância genética (CV_g) e a relação entre o coeficiente de variação genético e experimental (CV_r) foram calculados pelas equações $CV_g = (\hat{\sigma}_g / \hat{\mu}) \times 100$ e $CV_r = \frac{(\hat{\sigma}_g / \hat{\mu}) \times 100}{(\hat{\sigma}_e / \hat{\mu}) \times 100}$, em que $\hat{\sigma}_g^2$ é a estimativa do valor genético, $\hat{\sigma}_e^2$ é a estimativa do valor experimental, e $\hat{\mu}$ é a média.

Os valores preditos foram usados para o cálculo da Distância Euclidiana Média Padronizada (DEMP) entre os genótipos. O cálculo foi feito usando a função “dist” após padronização dos dados pela função “scale” do software R. A contribuição relativa dos caracteres para a divergência genética (CCD) foi calculada pelo método proposto por Singh (1981) usando a função “singh” do pacote *biotools* no software R.

O dendrograma foi calculado com base na matriz de distância euclidiana pelo método UPGMA usando a função “hclust” e plotado com a função “plot” do software R. Para a determinação do número de grupos formados foi usado o critério estatístico de Mojena

(MOJENA, 1977), calculado no software Genes (CRUZ, 2013). O ponto de corte foi então inserido no dendrograma com a função “abline” do software R.

2.5.1.3 Caracteres Qualitativos

Para os caracteres qualitativos, foi realizada a contagem do número e calculada a frequência dos genótipos por classe de cada caractere.

Uma matriz das classes dos caracteres qualitativos (colunas) *versus* os genótipos, contendo os números representativos das classes de cada caractere foi construída e usada para o cálculo da dissimilaridade entre os genótipos. A dissimilaridade foi calculada pelo complemento aritmético do índice de coincidência simples (CRUZ; CARNEIRO; REGAZZI, 2014) usando a função “Distancia” do pacote *MultivariateAnalysis*.

A matriz de dissimilaridade calculada foi usada para gerar o dendrograma UPGMA com as funções “hclust” e “plot” do software R. Para a determinação do número de grupos formados foi usado o critério estatístico de Mojena (MOJENA, 1977), calculado no software Genes (CRUZ,). O ponto de corte foi inserido no dendrograma com a função “abline” do software R.

2.5.2 Correlação entre caracteres

A correlação foi avaliada entre os caracteres quantitativos, entre os qualitativos e entre as matrizes de distância ou dissimilaridade dos caracteres quantitativos, qualitativos e marcadores SNPs (Figura 2).

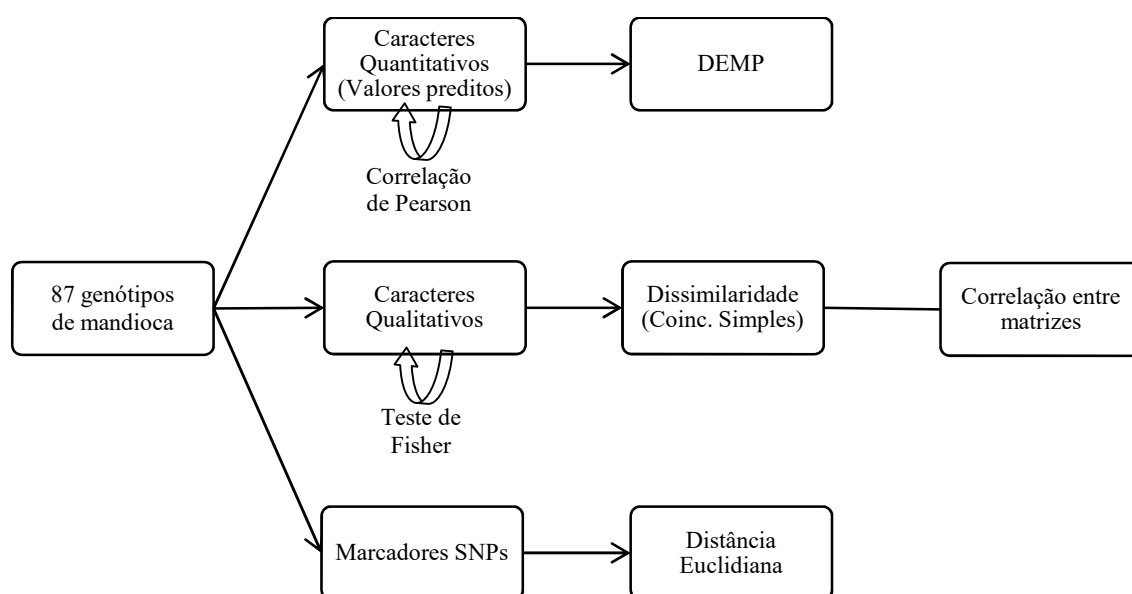


Figura 2. Workflow das análises de correlação de 87 genótipos de mandioca com caracteres quantitativos e qualitativos e marcadores SNPs.

A correlação entre os caracteres quantitativos foi calculada pelo coeficiente de correlação de Pearson usando a função “chart.Correlation” do pacote *PerformanceAnalytics*.

A associação entre os caracteres qualitativos foi calculada pelo teste de Fisher usando a função “fisher.test”. As associações significativas foram plotadas em gráficos de barras empilhadas usando a função “ggbarstats” do pacote *ggplot2*.

Para verificar a existência de correlação entre os caracteres quantitativos, qualitativos e moleculares, foi avaliada a correlação entre as matrizes de distância de todos os caracteres. Para tanto, foi calculada uma matriz de distância com cada caractere quantitativo e qualitativo e uma com os SNPs. A correlação foi calculada pelo teste de Mantel (MANTEL, 1967) usando a função “mantel” do pacote *vegan*.

Quando as correlações entre as matrizes de distância dos caracteres quantitativos e qualitativos foram significativas, foram plotados boxplots dos valores das características quantitativas por classe das qualitativas usando a função “boxplot”.

Quando as correlações entre a matriz de distância dos SNPs e as dos caracteres quantitativos e qualitativos foram significativas, buscou-se identificar se haviam amostras de SNPs com alta correlação com as características da seguinte forma (Figura 3): foram geradas amostras com 50; 100; 200; 300; 500; 750; 1000; 1500; 2000; 3000; 4000; 10000 SNPs, com 100 réplicas de reamostragem. Com estas amostras, foram geradas matrizes de distância euclidiana entre os 87 genótipos, as quais foram comparadas com a matriz gerada com a característica pelo teste de correlação de Mantel (1967). Um gráfico de dispersão do número de SNPs *versus* o valor de correlação foi plotado e foi identificado o número de SNPs a partir do qual não havia aumento expressivo no valor da correlação (n SNPs). Gerou-se então 10.000 amostras de n SNPs e calculou-se a matriz de distância de cada uma delas. Posteriormente, foi verificada a correlação destas matrizes com a característica. As amostras foram ordenadas de forma decrescente em relação aos valores de correlação e foram selecionadas 10% das amostras com as maiores correlações. Dentre estas, foram selecionados os SNPs mais frequentes. Foi então plotado um gráfico de dispersão dos valores de correlação dos SNPs mais frequentes *versus* o número de SNPs e, novamente, foi identificado o número de SNPs a partir do qual não havia aumento expressivo no valor da correlação (x SNPs). Dos x SNPs, foram extraídas 5000 subamostras de 4; 15 e 50 SNPs e a correlação das matrizes destas amostras com a matriz da característica foram calculadas. Foram selecionadas as amostras com maiores correlações. Os valores de 4; 15 e 50 SNPs foram determinados por análise visual do gráfico dos valores de correlação dos SNPs mais frequentes *versus* o número

de SNPs. Foram gerados *heatmaps* e boxplots dos quatro SNPs com as maiores correlações usando a função “pheatmap” do pacote *pheatmap* e a função “boxplot”, respectivamente.

Todas as análises foram realizadas no software R.

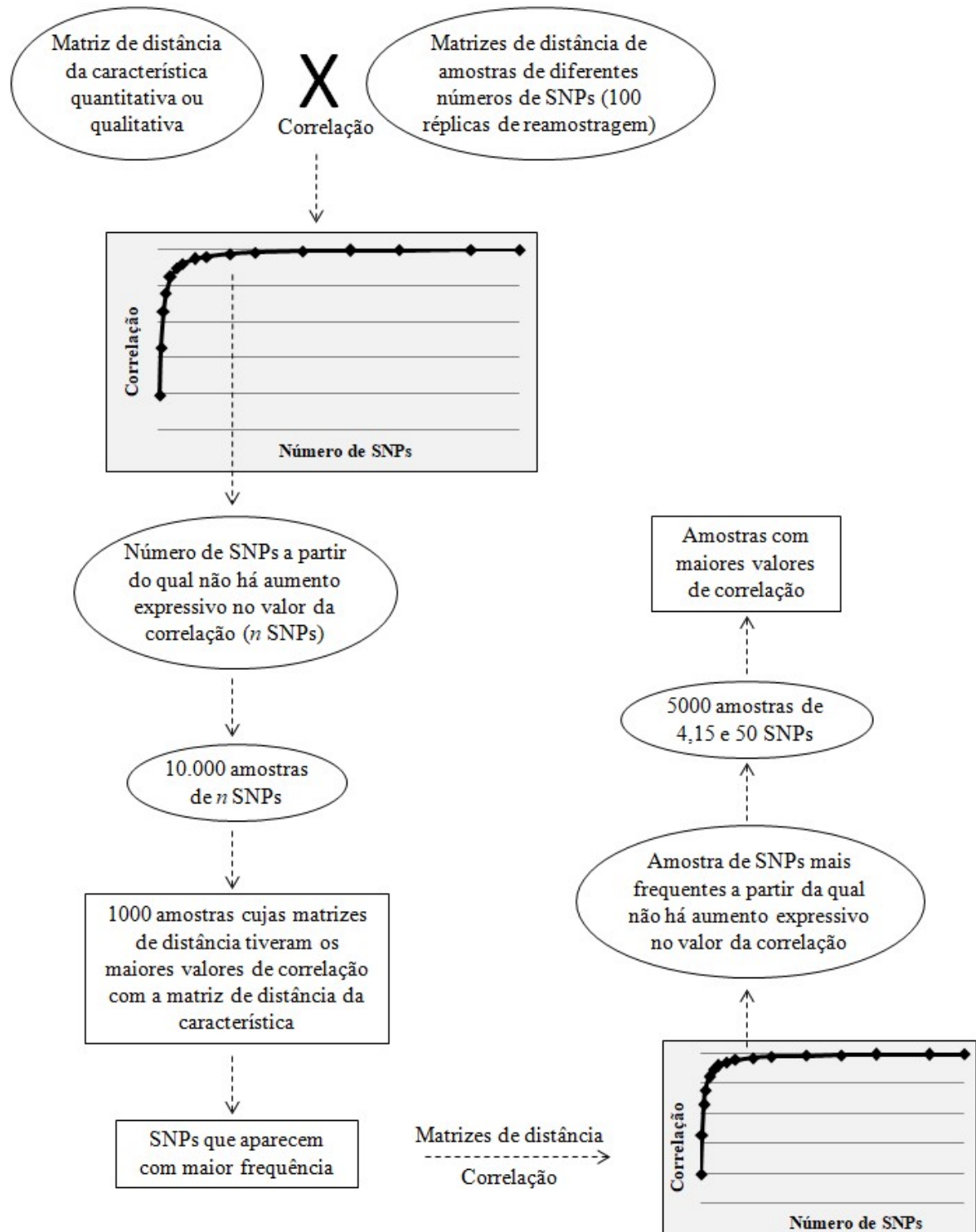


Figura 3. *Workflow* para identificação de amostras de SNPs com alta correlação com características quantitativas e qualitativas.

3 RESULTADOS

3.1 DIVERSIDADE GENÉTICA POR MARCADORES MOLECULARES

Foram identificados 388.349 SNPs no genoma de *Manihot esculenta* e 17.816 SNPs (4,59%) foram selecionados após a exclusão dos marcadores com MAF menor que 0,05. O número de alelos foi igual a dois para todos os SNPs e as frequências alélicas médias foram de 0,74 para o alelo mais comum e de 0,26 para o alelo menos comum.

A heterozigosidade observada (H_o) por genótipo variou de 0,39 a 0,53, com média igual a 0,46; a heterozigosidade esperada (H_e) variou de 0,1 a 0,5, com média de 0,35; e o índice de fixação (F_i) variou de -0,53 a -0,13, com média de -0,34.

A H_o por SNP variou de 0,01 a 1,0, com 41% dos marcadores com valores maiores que 0,5. O conteúdo de informação polimórfica (PIC) variou de 0,09 a 0,38, com média de 0,28 e 64% dos SNPs com PIC superior a 0,25. Foram identificados desvios do equilíbrio de Hardy-Weinberg (p -valor $< 0,01$) em 59% dos SNPs.

Os genótipos se apresentaram molecularmente diversificados, o que foi evidenciado pela matriz de distância e pela análise de agrupamento. Os valores de distância euclidiana variaram de 0,49 (G11 e G76) a 0,82 (G44 a G118), com média de 0,68 e aproximadamente 80% dos valores entre 0,65 e 0,75 (Figura 4A). O G118 foi o genótipo que apresentou a maior média de distância euclidiana em relação aos demais, de 0,75.

No dendrograma UPGMA (Figura 5A) é possível observar que há alta divergência entre todos os genótipos, representada pelos longos ramos individuais e pelo alto número de grupos formados – 17, de acordo com critério de Mojena (1977) – sendo sete grupos constituídos por apenas um genótipo (G30, G118, G115, G26, G120, G7 e T2).

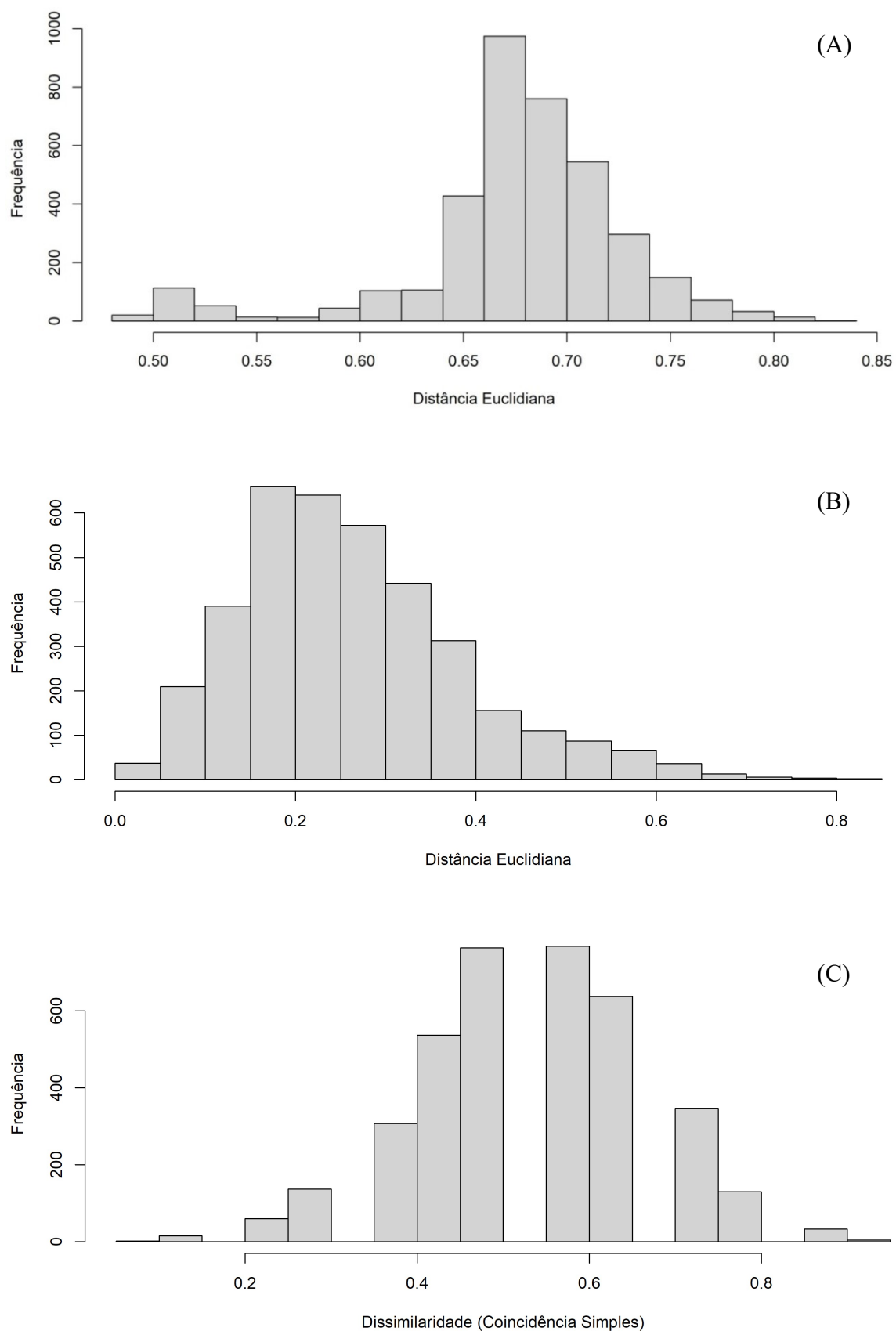


Figura 4. Distribuição dos valores de distância genética entre 87 genótipos de mandioca com base em marcadores SNPs (A) e caracteres quantitativos (B) e qualitativos (C).

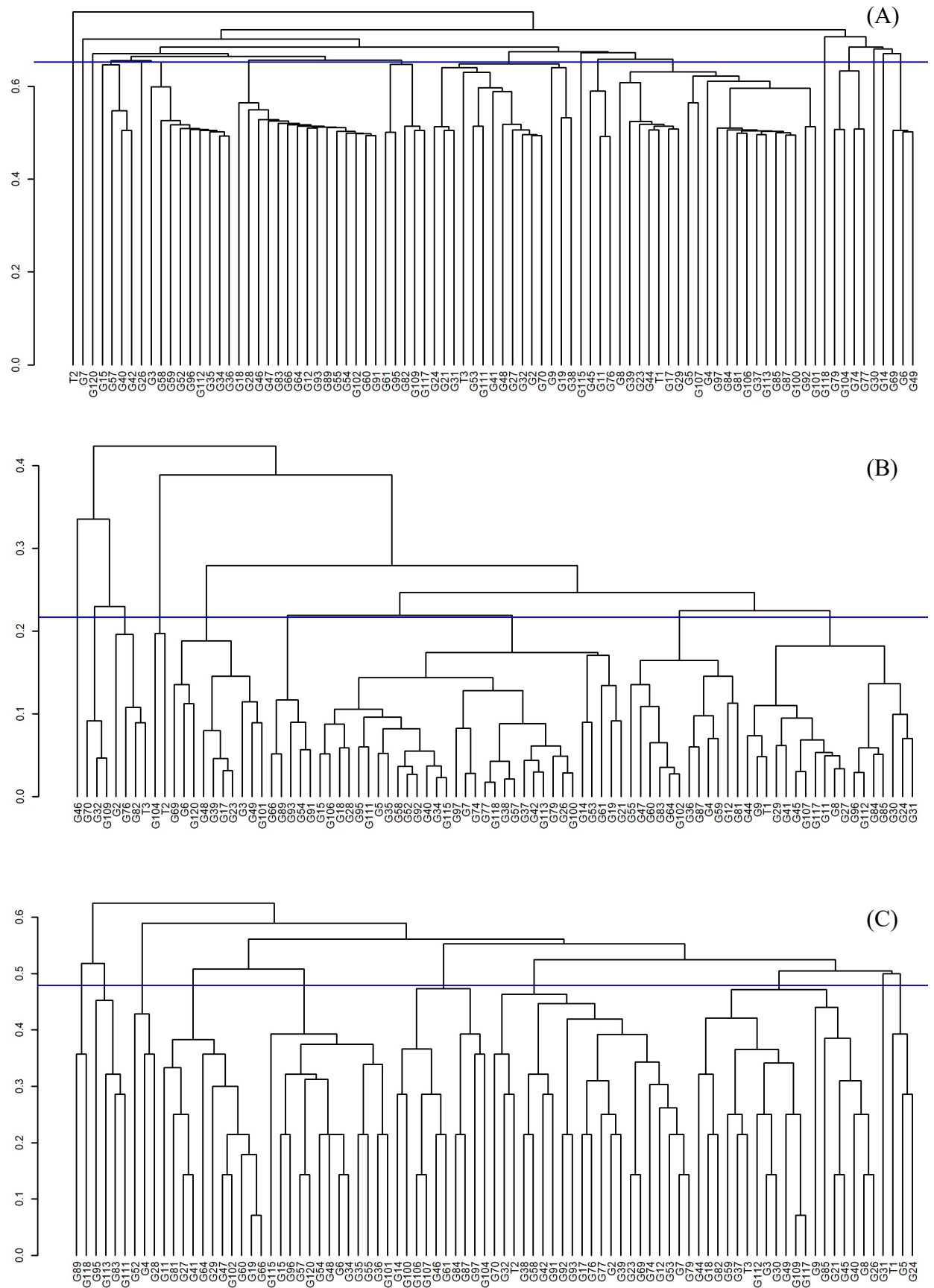


Figura 5. Dendrograma UPGMA de 87 genótipos de mandioca com base em marcadores SNPs (A) e caracteres quantitativos (B) e qualitativos (C).

3.2 DIVERSIDADE GENÉTICA POR CARACTERES QUANTITATIVOS

Foram avaliados os caracteres Altura da Parte Aérea (HPA), Número de Raízes Totais (NRT), Peso de Raízes Totais (PESORT) e Peso de Raízes Comerciais (PESORC) dos 87 genótipos de mandioca. A HPA variou de 2,29 (G69) a 2,96 m (G46), com média de 2,59; o NRT, de 5,94 (T2) a 7,64 raízes (T3), com média de 7,00; o PESORT variou de 2,64 (T2) a 2,82 kg (G2), com média de 2,71; e o PESORC variou de 2,30 (G104) a 2,50 kg (G2), com média de 2,37.

Os valores supracitados foram preditos pela metodologia REML/BLUP, ou seja, são os valores genotípicos livres do efeito ambiental, desta forma, os genótipos podem ter apresentado valores maiores ou menores em campo de acordo com a influência do ambiente. Portanto, é possível comparar os valores dos genótipos entre si sem a realização de testes, como testes de média, pois já foi excluída a variância ambiental.

O T3 destacou-se por estar entre os três genótipos de maiores valores para as quatro características. O G82 também apresentou altos valores de todos os caracteres. Os G2, G76, G109, G70 e G32 apresentaram altos valores de PESORT e PESORC, podendo ser melhorados quanto ao NRT e/ou HPA, o que poderia elevar ainda mais a produção.

A herdabilidade foi alta para o caractere HPA, de 0,79 e moderada para os demais caracteres, de 0,55, 0,62 e 0,63 para NRT, PESORT e PESORC.

A distância euclidiana média padronizada entre os genótipos variou de 0,02 (G56 e G83) a 0,85 (G74 e G87), com uma média geral de 0,26 e 80% dos valores entre 0,1 e 0,4 (Figura 4B). O T3 foi o genótipo que apresentou a maior média de distância euclidiana em relação aos demais, de 0,50. Foram formados nove grupos entre os genótipos pela análise de agrupamento UPGMA (Figura 5B), sendo o maior grupo formado por 32 genótipos e o menor grupo formado por apenas um genótipo, o G46. O NRT apresentou o maior valor de contribuição relativa dos caracteres para divergência (CCD), de 84%, seguido pelo HPA (15%), PESORT (0,6%) e PESORC (0,4%).

3.3 DIVERSIDADE GENÉTICA POR CARACTERES QUALITATIVOS

Dos 14 caracteres qualitativos avaliados nos 87 genótipos de mandioca, a forma do lóbulo central foi a característica com maior número de classes, seguida da cor externa do caule e da cor do pecíolo (Tabela 3).

O cozimento da raiz de 95% dos genótipos demorou entre zero e 20 minutos. Três genótipos (G15, G93 e T1) tiveram o cozimento entre 21 e 30 minutos e o genótipo T2 teve o cozimento após 30 minutos. A cor externa da raiz foi predominantemente marrom, com apenas dois genótipos (G4 e G28) com cor branca ou creme. A cor predominante do córtex da raiz foi branca ou creme, seguida pela cor rosada e amarela e, apenas sete genótipos apresentaram a cor roxa. A cor da polpa rosada foi rara entre os genótipos, notada apenas no genótipo 118. A cor amarela da polpa também foi pouco encontrada - em apenas sete genótipos; e as cores branca e creme foram as mais comuns, observadas em 63 e 28% dos genótipos, respectivamente. As raízes foram predominantemente rugosas (93%) e sem pelos (84%), com apenas seis genótipos com raízes de textura lisa e 14 genótipos com presença de pelos nas raízes. O pedúnculo das raízes não estava presente em 33% dos genótipos; 45% tinham pedúnculo em todas as raízes e 22% dos genótipos eram mistos, ou seja, havia raízes com e sem pedúnculo na mesma planta.

Foram caracterizadas oito classes da forma do lóbulo central. A forma elíptica foi a mais presente (41%), seguida da lanceolada (27%) e da reta (24%). As formas obovada, oblonga, piramidal, pandurada e linear-hostilobada foram pouco frequentes. As duas primeiras ocorreram em dois genótipos cada; e as três últimas, em apenas um genótipo cada. A cor da folha apical foi predominantemente verde (46% verde-claro e 48% verde-arroxeadado) e apenas cinco genótipos apresentaram a cor roxa. A cor do pecíolo foi predominantemente vermelha (44%), mas também houve pecíolos com outras cinco cores: verde-avermelhado (18%), roxo (15%), verde-amarelado (8%), verde (8%) e vermelho-esverdeado (7%). A cor do córtex do caule variou entre verde-claro e verde-escuro, com frequência de aproximadamente 50% para cada cor. A cor externa do caule variou entre sete classes, sendo a cor prateada a mais observada (37%) e as cores verde-amarelada e laranja, as menos presentes, notadas em apenas três e dois genótipos, respectivamente. A maioria dos genótipos (74%) não apresentou doença foliar e a floração esteve presente em 56% dos genótipos e ausente nos demais 44%.

Tabela 3. Número e frequência de genótipos por classe de 14 caracteres qualitativos avaliados em 87 genótipos de mandioca.

Descritor / Classes	Nº de Genótipos	Frequência (%)	Descritor / Classes	Nº de Genótipos	Frequência (%)
<i>Cozimento (Coz)</i>			<i>Forma do Lóbulo Central (ForLC)</i>		
< 10 minutos	44	50.57	elíptica	36	41.38
11 a 20 minutos	39	44.83	lanceolada	23	26.44
21 a 30 minutos	3	3.45	reta	21	24.14
> 30 minutos	1	1.15	obovada	2	2.30
<i>Cor Externa da Raiz (CorER)</i>			piramidal	1	1.15
marrom claro	41	47.13	pandurada	1	1.15
marrom escuro	44	50.57	oblonga	2	2.30
branco ou creme	2	2.30	linear-hostilobada	1	1.15
<i>Cor do Córtex da Raiz (CorCR)</i>			<i>Cor do Pecíolo (CorPe)</i>		
branco ou creme	40	45.98	verde-amarelado	7	8.05
rosado	27	31.03	vermelho	38	43.68
amarelo	13	14.94	roxo	13	14.94
roxo	7	8.05	verde-avermelhado	16	18.39
<i>Cor da Polpa (CorP)</i>			vermelho-esverdeado	6	6.90
branca	55	63.22	verde	7	8.05
rosada	1	1.15	<i>Cor do Córtex do Caule (CorC)</i>		
creme	24	27.59	verde-claro	43	49.43
amarela	7	8.05	verde-escuro	44	50.57
<i>Textura da Epiderme da Raiz (TextEP)</i>			<i>Cor Externa do Caule (CorEC)</i>		
lisa	6	6.90	marrom-claro	6	6.90
rugosa	81	93.10	cinza	17	19.54
<i>Pubescência (Pub)</i>			prateado	32	36.78
ausente	73	83.91	marrom-escuro	12	13.79
presente	14	16.09	dourado	15	17.24
<i>Presença de Pedúnculo nas Raízes (PrePed)</i>			laranja	2	2.30
séssil	29	33.33	verde-amarelado	3	3.45
pedunculada	39	44.83	<i>Floração (Flor)</i>		
mista	19	21.84	ausente	38	43.68
<i>Cor da Folha Apical (CorFA)</i>			presente	49	56.32
verde-claro	40	45.98	<i>Doença Foliar (DoecFol)</i>		
roxo	5	5.75	ausente	64	73.56
verde-arroxado	42	48.28	presente	23	26.44

A matriz de dissimilaridade calculada pelo complemento aritmético do índice de coincidência simples e a análise de agrupamento UPGMA evidenciaram que é possível diferenciar todos os 87 genótipos com as características avaliadas. A distância variou de 0,07 (G4 e G113) a 0,93 (G19 e G66), com média de 0,54 (Figura 4C). A análise de agrupamento permitiu a formação de 10 grupos (Figura 5C), sendo um grupo formado por apenas um genótipo, o G31.

3.4 CORRELAÇÃO ENTRE CARACTERES QUANTITATIVOS

A correlação entre os caracteres quantitativos foi significativa e positiva para todos os pares de caracteres, com exceção de HPA *versus* NRT (Figura 6).

O maior valor de correlação foi entre PESORC e PESORT. Ambos os caracteres se referem aos pesos das raízes, porém, para a obtenção de PESORT foram consideradas todas as raízes e, para a obtenção de PESORC, foram avaliadas apenas as raízes adequadas para o comércio.

O valor de correlação de 0,50 entre NRT e PESORT sugere que, em aproximadamente metade dos casos, quanto maior o número de raízes, maior o peso total destas. O gráfico de dispersão mostrou que os dois genótipos que apresentaram os menores valores de NRT também apresentaram os menores valores de PESORC (T2 e G104) e que os genótipos com os maiores valores de PESORT também tinham altos valores de NRT (G2, T3 e G82). Porém também é possível perceber genótipos com altos valores de NRT e com valores de PESORT baixos a médios, enquanto genótipos com NRT baixo não apresentaram alto PESORT. Assim, é possível concluir que os genótipos com muitas raízes (acima de sete, aproximadamente) podem ser altamente produtivos ou não, mas os genótipos com poucas raízes (abaixo de sete) dificilmente atingirão altas produções, com pesos acima de 2,8 kg por planta. A correlação entre NRT e PESORC (0,33) foi menor que a observada entre NRT e PESORT, provavelmente porque as raízes não comerciais foram descartadas na avaliação de PESORC. Invariavelmente, também não foram observados altos valores de PESORC em genótipos com baixo NRT.

A HPA apresentou correlação de 0,37 com o PESORT e de 0,42 com o PESORC. Observando o gráfico de dispersão, é possível perceber que as plantas mais baixas tendem a produzir raízes com peso entre baixo e médio, enquanto nas plantas mais altas, a variação do peso das raízes tende a ser entre médio e alto. Essa correlação provavelmente está relacionada ao vigor das plantas, que permite um maior acúmulo de energia nas raízes e um maior engrossamento destas, já que a HPA não apresentou correlação com o número de raízes. Isto

também explica o fato de a correlação com PESORC ser maior que aquela com PESORT, já que as raízes com baixo comprimento e diâmetro não foram consideradas como comerciais.

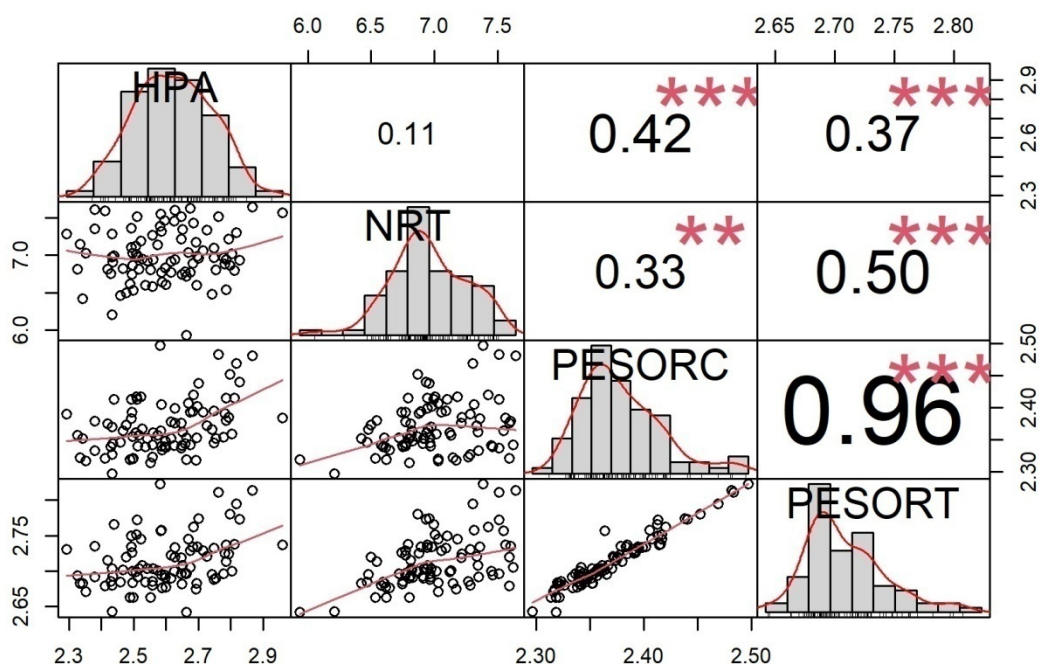


Figura 6. Correlação e Dispersão entre os caracteres quantitativos Altura de Plantas (HPA), Número de Raízes Totais (NRT), Peso de Raízes Comerciais (PESORC), e Peso de Raízes Totais (PESORT) medidos em 87 genótipos de mandioca.

3.5 ASSOCIAÇÃO ENTRE CARACTERES QUALITATIVOS

O teste de Fisher, realizado entre todos os pares de caracteres qualitativos, foi significativo entre TextEp e CorER; CorCR e CorP; CorCR e Pub; CorP e CorC; CorEC e CorC; e ForLC e CorC, considerando p-valor de 0,01. As quatro primeiras associações foram representadas em gráficos de barras empilhadas (Figura 7).

A TextEp foi rugosa em todos os genótipos com CorER marrom-escuro e em 90% dos genótipos com CorER marrom-clara e a textura foi lisa nos dois genótipos com CorER branca ou creme e em 10% dos genótipos com CorER marrom-clara (Figura 7A). Assim, existe uma tendência nesta população de as raízes com coloração externa mais escura apresentarem rugosidade e aquelas com coloração clara serem lisas.

As raízes com CorCR branca ou creme, rosada ou roxa tiveram a CorP majoritariamente branca, enquanto aquelas com CorCR amarela tiveram a CorP principalmente creme (11 genótipos com polpa creme e dois com polpa branca) (Figura 7B). A CorP amarela foi identificada predominantemente em raízes com CorCR rosada, sendo que, dos sete genótipos com polpa amarela, cinco tinham o córtex da raiz rosado, um tinha o córtex roxo e o outro, branco ou creme. O único genótipo com CorP rosada também tinha o córtex da raiz rosado.

A pubescência (Pub) foi mais presente em genótipos com CorCR roxa (57%), seguido por genótipos com CorCR rosada (26%), amarela (8%) e branca ou creme (5%) (Figura 7C), indicando que genótipos com córtex da raiz de coloração mais escura têm uma tendência maior de apresentarem pelos nas raízes.

A CorC verde-clara foi predominante nos genótipos com CorP amarela (100%) e creme (62%), enquanto a CorC verde-escura predominou nos genótipos com CorP branca (62%) (Figura 7D). O único genótipo com polpa rosada teve a CorC verde-escura.

Todos os genótipos com CorEC dourada, laranja ou verde-amarelada apresentaram a CorC verde-clara, enquanto os genótipos com CorEC cinza, marrom e prateada tiveram predominância da CorC verde-escura (gráfico não mostrado).

Os genótipos com CorC verde-clara apresentaram majoritariamente folhas com ForLC elíptica (56%) e lanceolada (28%) e apenas 9% tinham ForLC reta. Por outro lado, a maioria dos genótipos com CorC verde-escura tinham a ForLC reta (39%), seguida da forma elíptica (27%) e lanceolada (25%) (gráfico não mostrado).

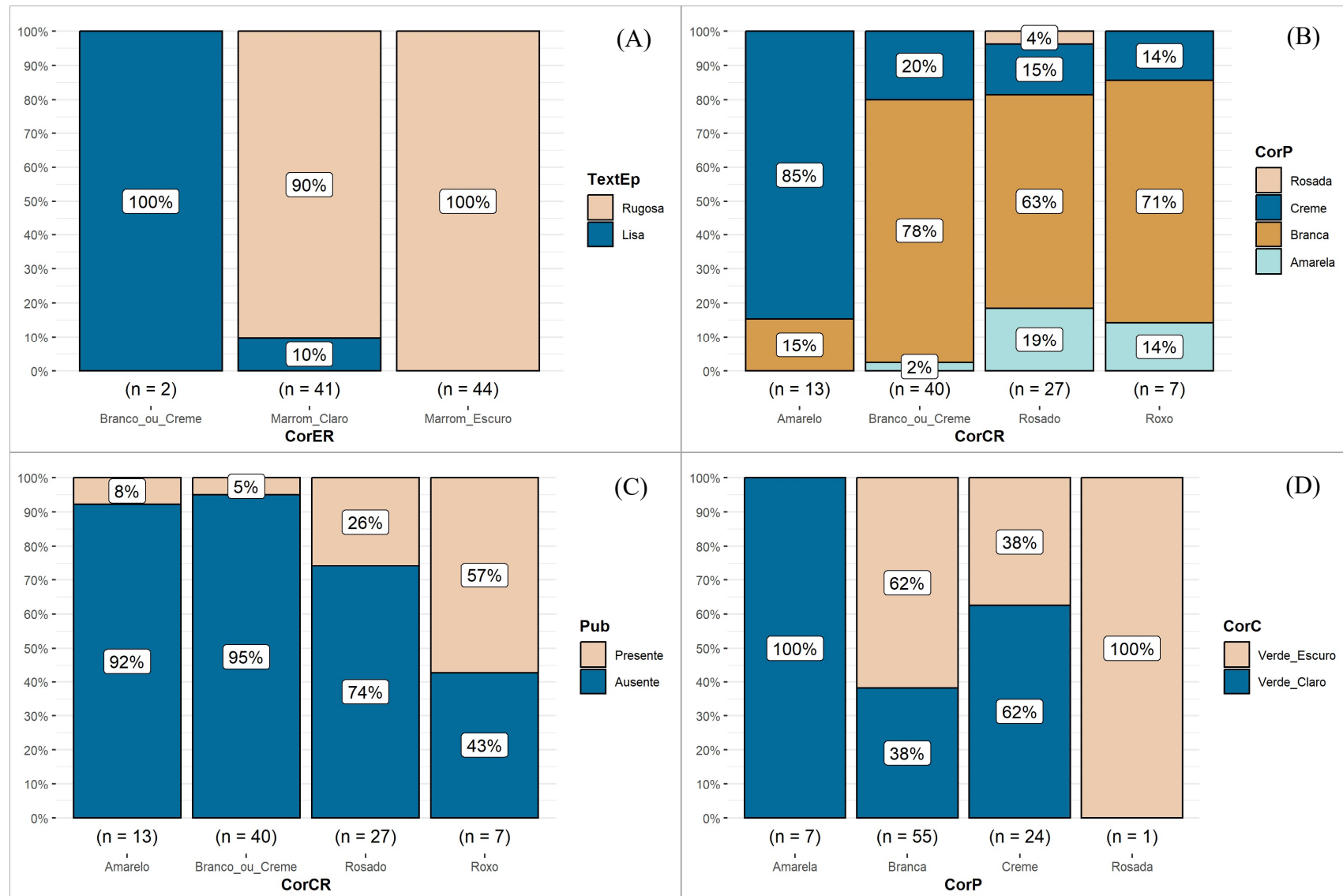


Figura 7. Gráfico de barras empilhadas de pares de caracteres qualitativos com associação significativa (p -valor $< 0,01$) pelo teste de Fisher. CorER = Cor Externa da Raiz; TextEp = Textura da Epiderme da Raiz; CorCR = Cor do Córtex da Raiz; CorP = Cor da Polpa da Raiz; Pub = Pubescência da raiz; CorC = Cor do Córtex do Caule. Os caracteres foram avaliados em 87 genótipos de mandioca.

3.6 CORRELAÇÃO ENTRE MATRIZES DE DISTÂNCIA DOS CARACTERES QUANTITATIVOS E QUALITATIVOS E SNPS

As correlações foram significativas entre as matrizes de distância de NRT e Coz; PESORC e CorFA; e PESORT e CorFA, considerando um p-valor de 0,01, entretanto os valores de correlação foram baixos, de 0,17; 0,08; e 0,05, respectivamente. Para verificar a origem das correlações entre os caracteres, foram plotados boxplots dos valores das características quantitativas por classe das qualitativas (Figura 8), considerando um intervalo de confiança de 95%, representado pelos entalhes nas caixas dos boxplots.

O boxplot do número de raízes por classe de tempo de cozimento (Figura 8A) demonstrou que apenas a classe >30 apresentou diferença das demais quanto ao número de raízes, tendo o menor NRT. Porém, como esta classe foi representada por apenas um genótipo, o T2, não é possível tirar conclusões a respeito da correlação entre os dois fatores. Serão necessários mais estudos para associar números de raízes muito baixos e demora no cozimento.

Em relação à correlação dos caracteres PESORT e PESORC com a CorFA (Figuras 8B e 8C), observou-se que a classe de CorFA verde-clara apresentou pesos de raízes ligeiramente maiores que as demais, considerando o intervalo de confiança. Além disso, os sete genótipos com maiores PESORT e PESORC (G2, G82, T3, G76, G109, G32 e G70) apresentaram a cor da folha apical verde-clara.

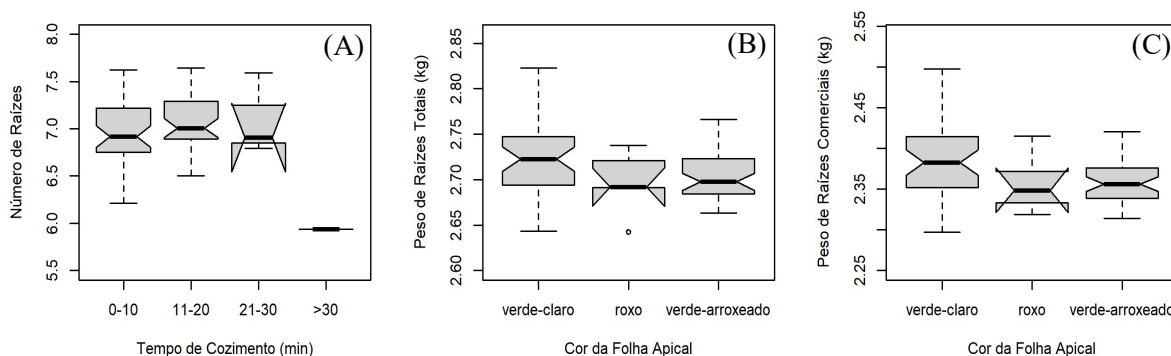


Figura 8. Boxplots dos valores das características quantitativas Número de Raízes Totais, Peso de Raízes Totais e Peso de Raízes Comerciais por classe das qualitativas Tempo de Cozimento ou Cor da Folha Apical observadas em 87 genótipos de mandioca. Os entalhes não sobrepostos dão cerca de 95% de confiança de que duas medianas diferem.

A matriz de distância gerada com os SNPs teve correlação significativa com aquela calculada com o número de raízes (NRT) e com o peso de raízes totais (PESORT).

Para identificar se haviam subamostras de SNPs com alta correlação com essas características, foram geradas 100 subamostras de diferentes números de SNPs, calculadas as matrizes de distância com elas e verificadas a correlação destas matrizes com a matriz de NRT e de PESORT. Observou-se

que, a partir de 2000 SNPs, não havia aumento expressivo no valor da correlação. Gerou-se então 10.000 amostras de 2000 SNPs e calculou-se a matriz de distância de cada uma delas. Posteriormente, foi verificada a correlação destas matrizes com a matriz de NRT e com a matriz de PESORT. As amostras foram ordenadas de forma decrescente em relação aos valores de correlação e foram selecionadas 10% das amostras com maiores correlações. Dentre estas, foram selecionados os SNPs mais frequentes. Foi então plotado um gráfico de dispersão dos valores de correlação dos SNPs mais frequentes pelo número de SNPs (Figura 9), onde é possível observar que o valor de correlação das matrizes de SNPs com a matriz de NRT variou de 0,34 a 0,55 (um e 241 SNPs, respectivamente) e começou a diminuir com 287 SNPs. Estes 287 SNPs foram usados para obtenção de 5000 subamostras de 4, 15 e 50 SNPs. Foi possível obter conjuntos de 4; 15 e 50 SNPs cujas matrizes de distância apresentaram correlação com a matriz de NRT de 0,45; 0,50 e 0,56, respectivamente.

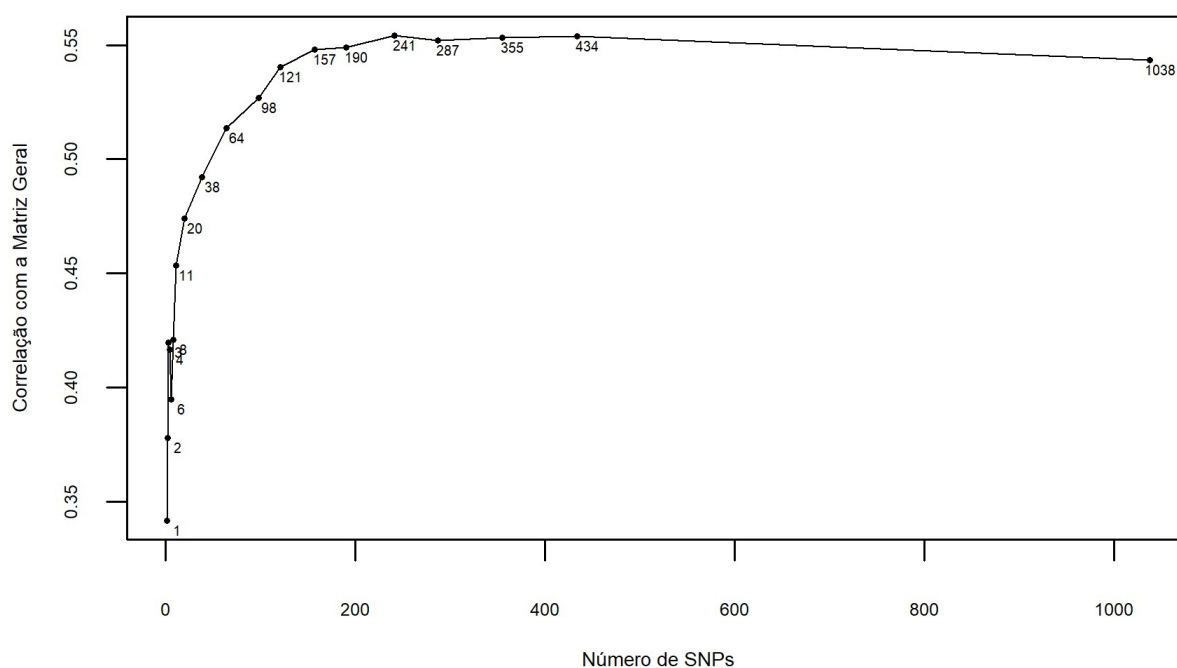


Figura 9. Gráfico de dispersão do número de SNPs *versus* os valores de correlação dos SNPs mais frequentes dentre 1000 amostras de 2000 SNPs cujas matrizes de distância tiveram as maiores correlações com a matriz de distância do Número de Raízes Totais de 87 genótipos de mandioca.

O conjunto com quatro SNPs cuja matriz de distância apresentou correlação com a matriz de NRT de 0,45 foi composto pelos SNPs S15_6957631 (S1), S11_5648439 (S2), S8_3549222 (S3) e S7_24722725 (S4) e permitiu diferenciar os genótipos quanto aos valores de NRT (Figura 10). Para fins de comparação, os valores de NRT foram divididos em quatro classes com base na média e no desvio padrão: baixo (<6,65), médio-baixo (6,65-7,00), médio-alto (7,00-7,35) e alto (>7,35).

O S1 foi o SNP mais frequente na primeira análise (Figura 10) e a matriz de distância gerada apenas com este SNPs teve correlação com a matriz do NRT de 0,34. Todos os 15 genótipos homozigotos para o alelo de referência nesse SNP (Grupos III, IV e V) apresentaram NRT alto ou médio-alto, com média de 7,35 raízes, variando de 7,03 a 7,62. Onze desses genótipos tiveram os S2, S3 e S4 homozigotos recessivos e foram os que apresentaram os maiores valores de NRT dentre os genótipos homozigotos para o alelo de referência, acima de 7,27 (Grupo IV), com média de 7,35. Os outros quatro genótipos (Grupos III e V) tiveram um ou dois heterozigotos dentre os S2, S3 e S4 e tiveram os menores valores dentre os genótipos homozigotos para o alelo de referência, variando de 7,03 a 7,17 e média de 7,19.

Dos 14 genótipos com NRT mais baixo (5,93- 6,94), 12 apresentaram S1 homozigoto para o alelo alternativo e dos 22 genótipos com S1 homozigoto para o alelo alternativo, 17 tiveram NRT baixo ou médio-baixo, o que sugere que genótipos homozigotos para o alelo alternativo tendem a apresentar valores baixos de NRT. Esta relação parece ser também dependente dos demais SNPs, pois os outros cinco genótipos com S1 homozigoto para o alelo alternativo apresentaram valores de NRT altos, mas diferenciaram-se dos demais genótipos nas suas constituições de S2, S3 e S4: Os G69, 6 e 49 (Grupo XII) apresentaram S2 homozigoto para o alelo alternativo e S3 e S4 heterozigotos; o G3 (Grupo XV) teve todos os quatro SNPs homozigotos para o alelo alternativo; e o T3, que foi o genótipo com maior NRT, apresentou S2 homozigoto para o alelo alternativo, S3 heterozigoto e S4 homozigoto para o alelo de referência.

Os dois genótipos com os maiores NRT (T3 e G120) tiveram constituições únicas dos quatro SNPs, sendo iguais para S2, S3 e S4 e com S1 heterozigoto no G120 e homozigoto para o alelo alternativo no T3. Os 50 genótipos que apresentaram S1 heterozigoto tiveram a maioria dos valores de NRT médio-altos ou médio-baixos, havendo apenas quatro genótipos com valores altos e dois com valores baixos.

Para a característica PESORT, os valores de correlação dos SNPs mais frequentes variaram de 0,24 (um SNP) a 0,52 (142 SNPs) e começaram a diminuir com 191 SNPs. Estes 191 SNPs foram usados para obtenção de 5000 subamostras de 4, 15 e 50 SNPs. Foi possível obter conjuntos de 4; 15 e 50 SNPs cujas matrizes de distância apresentaram correlação com a matriz de PESORT de 0,48; 0,50 e 0,55, respectivamente.

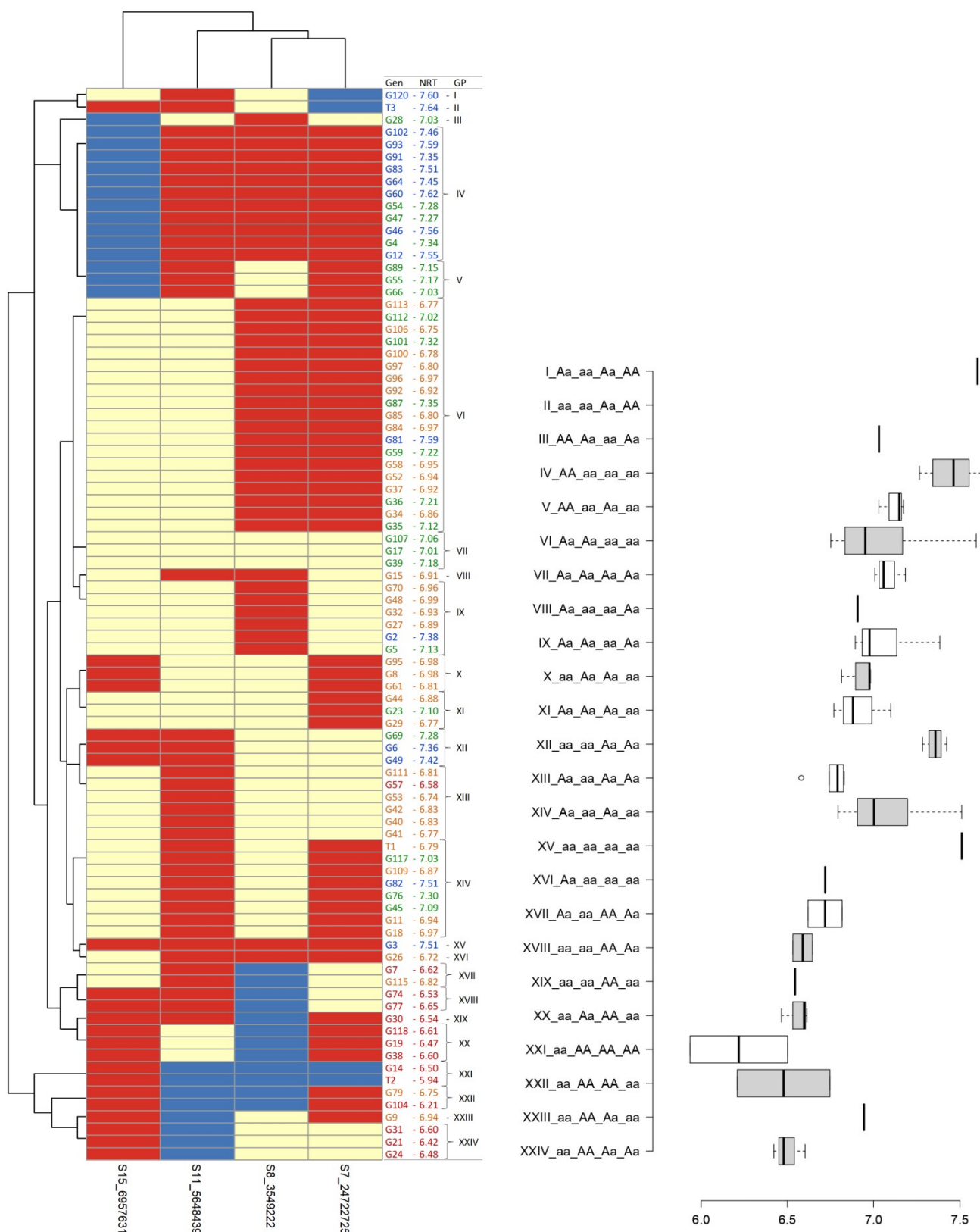


Figura 10. (A) Heatmap de quatro SNPs cujas matrizes de distância tiveram correlação de 0,45 com a matriz de distância do Número de Raízes Totais (NRT) de 87 genótipos de mandioca. Legenda do Heatmap: ■ homozigotos para o alelo de referência; ■ homozigotos para o alelo alternativo; ■ heterozigoto. Legenda dos genótipos e valores de NRT: ● <6,65; ● 6,65-7,00; ● 7,00-7,35, ● >7,35. (B) Boxplot do NRT por grupo de combinação de SNPs. A = alelo de referência; a = alelo alternativo.

Os quatro SNPs (S7_644347, S2_3333213, S2_4225917, S9_23440000) cujas matrizes de distância apresentaram correlação com a matriz de PESORT de 0,48 permitiram diferenciar a maior parte dos genótipos quanto aos valores de PESORT (Apêndice A). Quando os SNPs S2_3333213 e S2_4225917 são homozigotos para o alelo de referência (15 genótipos), os genótipos têm PESORT variando entre alto e médio-alto, com média de 2,76. Entretanto, quando os dois SNPs são homozigotos para o alelo alternativo (41 genótipos), 90% dos genótipos têm PESORT variando baixo ou médio-baixo, com média de 2,69. Na Figura 11 é possível observar a diferença dos valores de PESORT entre os genótipos agrupados de acordo com os SNPs S2_3333213 e S2_4225917.

Os quatro genótipos que apresentaram o S2_4225917 homozigoto para o alelo de referência juntamente com o S7_644347 homozigoto para o alelo de referência tiveram valores altos de PESORT, com média de 2,79, incluindo os dois genótipos com maiores PESORT, o T3 e o G2. Os dois genótipos com os menores valores de PESORT, T2 e G104, apresentaram, além dos S2_3333213 e S2_4225917 homozigotos para o alelo alternativo, o S9_23440000 homozigoto para o alelo de referência.

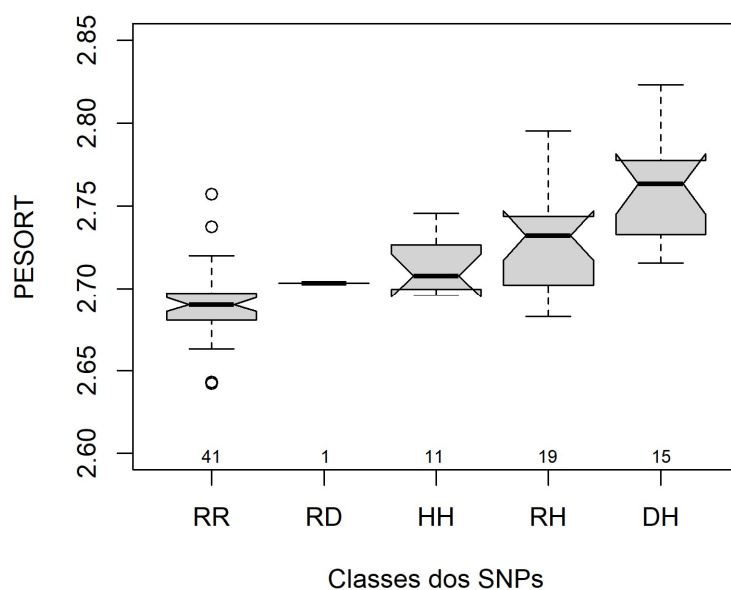


Figura 11. Boxplot dos valores de Peso de Raízes Totais (PESORT) entre os genótipos agrupados de acordo com os SNPs S2_3333213 e S2_4225917 em 87 genótipos de mandioca. RR = homozigotos para o alelo alternativo nos dois SNPs; RD = um homozigoto para o alelo alternativo e um homozigoto para o alelo de referência; HH = heterozigotos para os dois SNPs; RH = um homozigoto para o alelo alternativo e um heterozigoto; DH = um homozigoto para o alelo de referência e um heterozigoto. Os entalhes não sobrepostos dão cerca de 95% de confiança de que duas medianas diferem.

4 DISCUSSÃO

4.1 DIVERSIDADE GENÉTICA

A identificação e incorporação da diversidade genética representam o sucesso do melhoramento de plantas (SWARUP et al., 2021). Para que esse processo seja eficiente, é fundamental a avaliação da diversidade por diferentes tipos de caracteres e marcadores, como os morfoagronômicos e moleculares, de forma a caracterizar a variação fenotípica das características de interesse e os polimorfismos existentes no DNA, sem influência do ambiente (BHANDARI et al., 2017). Neste trabalho, foi analisada a diversidade genética de uma população de mandioca constituída de 87 genótipos cultivados por pequenos produtores no estado do Espírito Santo através de caracteres quantitativos, qualitativos e marcadores moleculares do tipo SNP.

Foram genotipados 17.816 SNPs de alta qualidade no genoma dos 87 genótipos de mandioca analisados. O número de SNPs foi similar ao relatado por Albuquerque et al. (2018), de 20.601, ao avaliarem um número consideravelmente maior de acessos, de 1580; também foi similar ao descrito por Pierre et al. (2021), que avaliaram 118 genótipos e detectaram 18 mil SNPs; e foi maior que o número genotipado por Karim et al. (2020), de 5.600 SNPs em 102 acessos. O número de SNPs pode variar de acordo com a população avaliada e com os métodos de detecção e filtragem usados (KUMAR; BANKS; CLOUTIER, 2012).

Os SNPs genotipados permitiram identificar ampla diversidade genética na população analisada, caracterizada por altas estimativas de H_e e H_o , distância genética e número de grupos de genótipos, bem como valores negativos de F_i . O valor médio de H_o (0,46) da população foi maior que o relatado por outros autores, de 0,36 (FERGUSON et al., 2011), 0,32 (DE OLIVEIRA et al., 2014), 0,23 (ALBUQUERQUE et al., 2018); e 0,26 (PIERRE et al., 2021). A H_o é a proporção de indivíduos heterozigotos observados na população e quanto maior o número de heterozigotos, maior a variação alélica e a combinação de alelos e, portanto, maior a diversidade (SAMUELS et al., 2016).

A H_e foi menor que a H_o , com média de 0,35 e valor máximo de 0,5; e o F_i , que é uma medida da taxa de endogamia, foi negativo. A H_o maior que H_e e o F_i negativo indicam maior número de heterozigotos em relação ao esperado sob o equilíbrio de Hardy-Weinberg (EHW) (ABRAMOVIS; BRASS; TASSABEHJI, 2020), o que também explica os 59% de SNPs identificados com desvios do EHW. Estes desvios podem ser explicados pelo fato de a

mandioca não constituir uma população verdadeira em relação aos cruzamentos ao acaso, tendo a reprodução assexuada (clonal) como principal forma de propagação da espécie (ALBUQUERQUE et al., 2018). Além disso, análises do genoma da mandioca têm evidenciado que o nível de heterozigosidade da espécie está entre os mais altos encontrados em genomas vegetais sequenciados, o que se atribui à sua natureza alógama, ampla distribuição tropical e propagação clonal (WANG et al., 2014; BREDESON et al., 2016; HU et al., 2021). O valor máximo de H_e de 0,5 é esperado em marcadores predominantemente bialélicos, como é o caso dos SNPs (HAMBLIN; WARBURTON; BUCKLER, 2007).

A H_o foi maior que 0,5 em 41% dos SNPs e o PIC foi superior a 0,25 em 64% dos SNPs. O PIC quantifica o valor da informação de polimorfismo de um marcador ao considerar o número e a frequência relativa dos alelos (ROMERO-SEVERSON et al., 2001). Marcadores com PICs maiores que 0,25 são considerados os marcadores mais informativos (BOTSTEIN et al., 1980), com potencial para os estudos genéticos sobre variabilidade, associação com fenótipos e seleção assistida (ALBUQUERQUE et al., 2018). Análises de SNPs em 1.580 acessos de mandioca pertencentes ao germoplasma internacional conservado no Brasil revelaram uma endogamia alta, com média de 0,69 e apenas 4% dos SNPs com H_o maior que 0,5 (ALBUQUERQUE et al., 2018). Esses resultados evidenciam o potencial do estado do Espírito Santo, com sua característica de agricultura familiar, em manter alta diversidade genética e, portanto, recursos genéticos úteis ao melhoramento. A inclusão de genótipos como esses cultivados no estado nos bancos de germoplasma podem aumentar a diversidade pela incorporação de novos alelos.

A ampla diversidade genética dos genótipos, detectada pelos SNPs, foi também observada fenotipicamente nos caracteres quantitativos altura de parte aérea (HPA), número de raízes totais (NRT) e peso de raízes totais e comerciais (PESORT e PESORC). A herdabilidade (h^2) dos caracteres variou de moderada a alta. A herdabilidade é definida como a proporção da variação total de uma característica fenotípica dentro de uma população que é atribuível à variação genética (SCHMIDT et al., 2019); é considerada alta quando os valores são maiores que 0,7 e moderada quando os valores se situam entre 0,2 e 0,7 (RAFFA; THOMPSON, 2017). O NRT apresentou h^2 de 0,55. Para esta característica, tem sido relatadas herdabilidades de 0,91 (AVIJALA et al., 2015) 0,65 (SILVA et al., 2016); 0,41 (PEPRAH et al., 2020); 0,38 (YONIS et al., 2020); e 0,32 (DE CARVALHO et al., 2022). A herdabilidade para o peso de raízes (PESORT e PESORC) foi de 0,62 e 0,64. Outros autores têm relatado herdabilidade variando de 0,33 a 0,78 para o peso de raízes totais (PEPRAH et al., 2020; YONIS et al., 2020). As características NRT e PESORT estão muito ligadas à

produtividade da mandioca. Para a produtividade de raízes, herdabilidades de 0,21; 0,35; 0,54; 0,87 têm sido relatadas (DE CARVALHO et al., 2022; ANDRADE et al., 2019; NTAWURUHUNGA; DIXON, 2010; AVIJALA et al., 2015, respectivamente). A altura da parte aérea foi o caractere com maior herdabilidade (0,79), o que evidencia que a altura das plantas está mais relacionada com a tendência genética dos genótipos em serem mais altos ou mais baixos.

Os valores moderados de herdabilidade apresentados pelas características NRT, PESORT e PESORC são desejáveis para o melhoramento da mandioca, pois indicam que os acessos que forem selecionados têm a tendência de transferir seus níveis de produtividade para as próximas gerações, mantendo, pelo menos, os mesmos patamares de produtividade (YOKOMIZO; FARIAS NETO, 2003)

Além da caracterização produtiva e molecular, foram analisados caracteres qualitativos, incluindo descritores morfológicos, tempo de cozimento da raiz e presença ou não de doenças. Dos descritores morfológicos, os mais variáveis foram pertencentes à parte aérea, a saber, a Forma do Lóbulo Central (ForLC), a Cor Externa do Caule (CorEC) e a Cor do Pecíolo (CorPe) com nove, sete e seis classes, respectivamente. Diaguna et al. (2022) observaram apenas três, seis e quatro classes das respectivas características supracitadas, provavelmente devido ao menor número de acessos analisados, de 29. Esses descritores estão entre os mais importantes para caracterização de variedades e da variabilidade morfológica da mandioca (SALES FILHO, 1991; SILVA et al., 2017)

As características de parte aérea, como são de pouco interesse econômico, são pouco visadas durante o processo de seleção, tanto aquela feita por produtores ou por programas de melhoramento, permitindo uma grande variabilidade entre as classes, mesmo em variedades melhoradas (TEIXEIRA et al., 2014).

Dentre as características da raiz, as mais variáveis foram a Cor do Córtex (CorCR) e a Cor da Polpa (CorP), com quatro classes cada, seguidas da Cor Externa (CorER) e a Presença de Pedúnculo (PrePed), com três classes cada. Esses caracteres são de grande importância para o melhoramento genético. A cor da polpa da raiz preferida pelos consumidores varia entre as regiões. Em determinados locais, há preferência por variedades de polpa amarela, enquanto em outros, somente as variedades de polpa branca ou creme são bem aceitas (PONTE, 2008). Na indústria, na maioria das regiões, a preferência é pela cor branca (FUKUDA et al., 2002; LORENZI, 2012). Os genótipos avaliados apresentaram polpa majoritariamente branca (63,2%) e creme (27,6%), mas também foram observados sete genótipos de coloração amarela e um de coloração roxa. Os onze genótipos com maiores

PESORT apresentaram raiz branca, sendo assim de grande interesse para o melhoramento da cultura.

A cor externa da raiz e a cor do córtex da raiz começaram a ter importância agrônômica há pouco tempo. Para a produção de mandioca de mesa são preferidas as mandiocas com coloração externa da raiz marrom-clara ou escura e para a indústria são visados acessos de cor do córtex e cor externa da raiz branca (VIEIRA et al., 2008). A CorER marrom foi observada em praticamente todos os genótipos, já a cor branca foi vista em apenas dois genótipos, os quais também apresentaram CorCR branca.

Em relação ao cozimento da raiz, praticamente todos os genótipos tiveram a raiz cozida com menos de 20 minutos, sendo que a raiz de metade dos genótipos cozinhou com menos de 10 minutos. Os genótipos T1, G15 e G93 foram os únicos a apresentar cozimento entre 21 e 30 minutos e o genótipo T2 foi o único demorar mais de 30 minutos. O genótipo T2 também apresentou os menores valores de NRT e PESORT, não sendo assim de interesse para seleção em programas de melhoramento para produtividade e cozimento, que são características de grande importância para os produtores e consumidores da raiz (MIRANDA et al., 2020). Os genótipos G15 e o G93 também apresentaram PESORT entre baixo e médio-baixo, já o genótipo T1 foi o décimo sétimo genótipo com maior PESORT e pode ser selecionado para produção, mas deverá ser melhorado quanto ao tempo de cozimento. Os genótipos com maiores PESORT apresentaram cozimento entre zero e 20 minutos, sendo que a maioria teve a raiz cozida entre 10 e 20 minutos.

As distâncias e dissimilaridades entre os genótipos foram calculadas pela distância euclidiana para os dados quantitativos e moleculares e pelo do complemento aritmético do índice de coincidência simples para os dados qualitativos e a média foi maior quando calculada com base nos marcadores moleculares. Isto provavelmente ocorre porque os marcadores SNPs são detectados em todo o genoma e em grande número, enquanto os caracteres quantitativos e qualitativos são fenótipos traduzidos de algumas regiões do DNA, além de serem poucos caracteres quando comparados com os quase 18 mil SNPs (PIERRE et al., 2022). Essa diferença também foi refletida no número de grupos formados pelo agrupamento UPGMA, sendo formados 17 grupos com os marcadores SNPs, nove com os caracteres quantitativos e dez com os qualitativos. A menor diversificação dos caracteres quantitativos em relação aos moleculares e qualitativos indica a possibilidade de seleção de genótipos similares com caracteres superiores, mas divergentes geneticamente.

Os agrupamentos foram, no geral, diferentes entre si, demonstrando a importância de se avaliar diferentes tipos de caracteres e marcadores para detectar a diversidade genética da

população (BHANDARI et al., 2017). Os caracteres quantitativos são fundamentais para os programas de melhoramento que visam produtividade, entretanto refletem pouco da diversidade dos genótipos e, sem a avaliação dos demais marcadores pode-se cometer o erro de se trabalhar com uma população de base genética estreita (COOPER et al., 2001). Os marcadores SNPs evidenciaram que este não é o caso da população avaliada, que apresentou ampla diversidade.

Os genótipos G70, G2, G32 e T3 mostraram similaridade nos agrupamentos por dados quantitativos e de SNPs, assim como os G77, G74, G104, G79, G118 e G14. Os primeiros genótipos estão entre os de maior PESORT e os últimos, entre os de menor NRT. Os genótipos G30, G118, G115, G26, G120, G7 e T2 formaram grupos unitários no dendrograma gerado com base nos SNPs. Dentre estes, destacam-se o G120, que é o genótipo com o terceiro maior NRT; o G118, que foi o único genótipo a apresentar polpa rosada; e o T2, que apresentou os mais baixos valores de NRT e PESORT e o pior tempo de cozimento dentre os genótipos estudados. O G120 e G118 podem ser estudados para o melhoramento das características NRT e cor da polpa rosada, respectivamente, contribuindo também para diversificar geneticamente as coleções de estudo dos programas de melhoramento. E o T2 pode ser incluído em estudos para identificação de marcadores relacionados à baixa produtividade e demora no cozimento.

4.2 CORRELAÇÃO ENTRE CARACTERES

Ocorreu correlação significativa entre PESORT *versus* PESORC, NRT e HPA; PESORC *versus* NRT e HPA; CorC *versus* CorP, CorEC e ForLC; CorCR *versus* Pub e CorP; CorER *versus* TextEP; Coz *versus* NRT; CorFA *versus* PESORT e PESORC; e SNPs *versus* NRT e PESORT.

Dentre os caracteres quantitativos avaliados, o PESORT e o PESORC são os de maior importância para o melhoramento visando alta produtividade de raízes, mas também são os de mais difícil avaliação, pois exigem o arranquio, limpeza e pesagem das raízes. A correlação ou associação destas características com outros caracteres ou marcadores moleculares pode facilitar e reduzir os custos dos programas de melhoramento ao permitir a aplicação de técnicas como a seleção indireta e a seleção assistida por marcadores moleculares.

Os caracteres PESORT e PESORC apresentaram correlação significativa e positiva com o NRT, a qual também tem sido relatada por outros autores (KARIM et al., 2020; PAZ et al., 2020; DIAGUNA et al., 2022). O NRT é uma característica um pouco mais fácil de avaliar

que o peso das raízes, pois exige apenas a contagem, não sendo necessária a limpeza e a pesagem, mas também necessita do arranquio. Porém, o NRT apresentou herdabilidade menor que PESORT e PESORC, o que torna a seleção indireta inviável.

A correlação de HPA com PESORT e PESORC, mesmo não sendo uma correlação alta, de 0,37 e 0,42, permitiu perceber que os genótipos com maiores pesos têm HPA entre médio e alto. Como HPA apresentou uma herdabilidade alta, maior que os demais caracteres e é uma variável mais fácil de avaliar, por estar acima do solo, pode ser uma alternativa para seleção indireta dos genótipos mais produtivos, eliminando-se os genótipos com HPA baixa. Entretanto, a correlação entre altura de plantas e peso de raízes caracterizada nesta população não tem sido observada em outras populações de mandioca, como as avaliadas por Vieira, Fialho e Carvalho, 2014, Karim et al., 2020, Paz et al., 2020 e Diaguna et al., 2022.

Por outro lado, a não correlação entre HPA e NRT e os valores baixos e médios de correlação com PESORT e PESORC podem sugerir potencial de aumentar ainda mais a produção através de cruzamentos, pois genótipos que apresentam uma característica alta podem ser cruzados com aqueles que apresentam apenas uma ou outra característica de valor elevado, visando à seleção de genótipos ainda mais produtivos. Um exemplo é o G60, que é o genótipo com segundo maior valor de NRT, mas teve pesos de raízes médio-baixos. Ou seja, o genótipo tem grande potencial de armazenamento, mas a produção de amido é baixa e há potencial de aumento na HPA. O cruzamento deste genótipo com genótipos que apresentam alto peso de raízes pode elevar ainda mais a produção. A manutenção da alta produtividade dos genótipos selecionados tem as vantagens de a cultura ser reproduzida vegetativamente e de as características apresentarem herdabilidade entre moderada e alta.

Outra característica da parte aérea com correlação com o PESORT e PESORC foi a CorFA. A correlação foi altamente significativa (p -valor = 0,002), porém de baixa magnitude. Mesmo assim, é uma correlação que deve ser mais bem estudada, pois os genótipos com maiores pesos apresentaram coloração da folha apical verde-clara e isto também pode ser observado no trabalho de Diaguna et al. (2022), onde, dentre 29 genótipos avaliados, os oito genótipos com maiores pesos também tinham CorFA verde-clara. Assim, essa associação, se confirmada, poderá contribuir para a identificação dos genótipos mais produtivos em conjunto com outras características e/ou com marcadores moleculares.

O PESORT também apresentou correlação com os SNPs, a qual foi identificada pela correlação positiva e significativa entre as matrizes de distância da característica com a dos SNPs. A correlação foi baixa quando foram considerados todos os SNPs, porém foi aumentada pela identificação de subconjuntos de SNPs pelas técnicas de reamostragem.

Contudo, foram caracterizados conjuntos de SNPs com correlação entre 0,48 e 0,55 com o PESORT, sendo que, com apenas dois SNPs foi possível classificar a maior parte dos genótipos como baixos e médio-baixos ou altos e médio-altos. Também foram encontrados SNPs correlacionados com o NRT, sendo que, com apenas quatro SNPs foi possível diferenciar os genótipos quanto aos valores de NRT. A associação destes SNPs com o NRT e com o PESORT deverá ser confirmada por análises de associação genômica ampla (GWAS) e em outras populações e, caso seja confirmada, poderão ser desenvolvidos marcadores para uso em seleção assistida.

A seleção assistida por marcadores acelera o melhoramento de culturas como a mandioca ao permitir a seleção de genótipos com características favoráveis em sementes, manivas, culturas de tecidos, plântulas, plantas jovens e/ou tecidos da parte aérea, sem a necessidade de esperar a produção de raízes tuberosas e de se realizar a colheita (KUMAWAT et al., 2020), cujo início ocorre entre 10 e 24 meses após o plantio (FIALHO et al., 2013).

5 CONCLUSÕES

A população de mandioca avaliada possui diversidade genômica, morfológica e produtiva suficiente para a seleção de genótipos com características agronômicas de interesse e para diversificação da cultura.

A avaliação com diferentes tipos de caracteres e marcadores permitiu obter resultados diversificados, apresentando diferentes aspectos da diversidade, complementando-se e caracterizando a população de forma abrangente.

Há correlações entre caracteres e marcadores, destacando-se a correlação de SNPs com as principais características de produtividade - número e peso de raízes - que poderão ser estudados para seleção assistida por marcadores.

6 REFERÊNCIAS

- ABRAMOV, N., BRASS, A., TASSABEHJI, M. Hardy-Weinberg Equilibrium in the Large Scale Genomic Sequencing Era. **Front. Genet.**, v. 11, n. 210. 2020. <https://doi.org/10.3389/fgene.2020.00210>.
- ALBUQUERQUE, H. Y. G.; CARMO, C. D.; BRITO, A. C.; OLIVEIRA, E. J. Genetic diversity of *Manihot esculenta* Crantz germplasm based on single-nucleotide polymorphism markers. **Ann. Appl. Biol.**, v. 173, p. 271–284, 2018. <https://doi.org/10.1111/aab.12460>
- ANDRADE, L. R. B. D.; SOUSA, M. B. E.; OLIVEIRA, E. J.; RESENDE, M. D. V. D.; AZEVEDO, C. F. Cassava yield traits predicted by genomic selection methods. **PLoS One**, v. 14, n. 11, 2018. <https://doi.org/10.1371/journal.pone.0224920>
- AVIJALA, M. F., BHERING, I. I., PEIXOTO, L. A., et al. Evaluation of cassava (*Manihot esculenta* Crantz) genotypes reveals great genetic variability and potential selection gain. **Australian J Crop Sci**, v. 9, n. 10, p. 940–947, 2015. <https://doi.org/10.3316/informit.659853541582368>
- BARBIERO, Natália Zardo. **Estudos genéticos em germoplasma de mandioca no estado do Espírito Santo, Brasil**. 2022. Tese (Doutorado em Agronomia) - Universidade Federal do Espírito Santo, Alegre, ES, 2022.
- BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v. 125, p. 375–385 (2020). <https://doi.org/10.1038/s41437-020-0312-1>
- BESTER, A. U.; CARVALHO, I. R.; SILVA, J. A. G. D.; et al. Three decades of cassava cultivation in Brazil: Potentialities and perspectives. **Revista Colombiana de Ciencias Hortícolas**, v. 15, n. 2, 2021. <https://doi.org/10.17584/rcch.2021v15i2.12087>
- BHANDARI, H. R., BHANU, A. N., SRIVASTAVA, K., et al. Assessment of genetic diversity in crop plants-an overview. **Adv. Plants Agric. Res.**, v. 7, n. 3, p. 279-286, 2017. <https://doi.org/10.15406/apar.2017.07.00255>
- BOTSTEIN, D.; WHITE, R. L.; SKOLNICK, M.; DAVIS, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **Am. J. Hum. Genet.**, v. 32, p. 314-331, 1980.
- BREDESON, J., LYONS, J., PROCHNIK, S. et al. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. **Nat. Biotechnol**, v. 34, p. 562–570, 2016. <https://doi.org/10.1038/nbt.3535>
- BYJU, G.; SUJA, G. Mineral nutrition of cassava. **Advances in Agronomy**, 159, 169-235, 2020. <https://doi.org/10.1016/bs.agron.2019.08.005>
- CARVALHO, M. S.; FERREIRA, M. F. D. S.; OLIVEIRA, W. B. D. S.; et al. Genetic diversity and population structure of *Euterpe edulis* by REML/BLUP analysis of fruit morphology and microsatellite markers. **Crop Breeding and Applied Biotechnology**, v. 20, n. 4, 2020. <https://doi.org/10.1590/1984-70332020v20n4a61>
- COOPER, H. D., SPILLANE, C., HODGKIN, T., COOPER, H. Broadening the genetic base of crops: an overview. In: COOPER, H. D., SPILLANE, C., HODGKIN, T. **Broadening the genetic base of crop production**, 2001, p. 1-23.

- CRUZ, C. D. Genes: a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, v. 35, p. 271-276, 2013. <https://doi.org/10.4025/actasciagron.v35i3.21251>
- CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. **Modelos Biométricos Aplicados ao Melhoramento Genético**. 3. ed. Viçosa: Editora UFV, 2014, 668 p.
- CRUZ, C. D.; FERREIRA, F. M.; PESSONI, L. A. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco: Suprema, 2011, 620 p.
- CULLIS, B. R.; SMITH, A. B.; COOMBES, N. E. On the design of early generation variety trials with correlated data. *Journal of agricultural, biological, and environmental statistics*, v. 11, n. 4, p. 381-393, 2006. <https://doi.org/10.1198/108571106X154443>
- DAR, A. A.; MAHAJAN, R.; SHARMA, S. Molecular markers for characterization and conservation of plant genetic resources. **Indian J Agric Sci**, v. 89, n. 11, 2019.
- DE CARVALHO, R. R. B., BANDEIRA E SOUSA, M., DE OLIVEIRA, L. A., & DE OLIVEIRA, E. J. Phenotypic diversity and selection in biofortified cassava germplasm for yield and quality root traits. **Euphytica**, v. 218, n. 12, 2022. <https://doi.org/10.1007/s10681-022-03125-6>
- DE OLIVEIRA, E. J., FERREIRA, C. F., DA SILVA SANTOS, V., et al. Potential of snp markers for the characterization of brazilian cassava germplasm. **Theor. Appl. Genet.**, v. 127, n. 6, p. 1423-40, 2014. <https://doi.org/10.1007/s00122-014-2309-8>
- DE OLIVEIRA, G. M.; SANTOS, J. D. O.; DOS SANTOS, C. B.; et al. Rise in temperature increases growth and yield of *Manihot* sp. plants. **Research, Society and Development**, v. 11, n. 9, 2022. <http://dx.doi.org/10.33448/rsd-v11i9.29891>
- DE SA, S. F.; DOS SANTOS, L. C. A.; DA CONCEIÇÃO, L. D. S.; et al. Genetic diversity via REML-BLUP of ex situ conserved macauba [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] ecotypes. **Genetic Resources and Crop Evolution**, v. 68, n. 8, p. 3193-3204, 2021. <https://doi.org/10.1007/s10722-021-01180-x>
- DINIZ, R. P.; OLIVEIRA, E. J. Genetic parameters, path analysis and indirect selection of agronomic traits of cassava germplasm. **Anais da Academia Brasileira de Ciências**, v. 91, n. 3, 2019. <https://doi.org/10.1590/0001-3765201920180387>
- FERGUSON, M. E., HEARNE, S. J., CLOSE, T. J., et al. Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. **Theoretical and Applied Genetics**, v. 124, n. 4, p. 685-95, 2011. <https://doi.org/10.1007/s00122-011-1739-9>
- FIALHO, J. D. F.; VIEIRA, E. A.; FIALHO, C. P. A. C.; et al. Mandioca no cerrado: orientações técnicas. Brasília, DF: Embrapa, 2013.
- FUKUDA, W. M. G.; FUKUDA, C.; DIAS, M. C.; et al. **Cultivares de mandioca recomendadas para o Brasil**. Cruz das Almas: Embrapa, 2003. 8p.
- FUKUDA, W. M. G.; IGLESIAS, C.; SILVA, S. O. Melhoramento de Mandioca. Cruz das Almas: Embrapa, 2003. 53p.
- HAMBLIN, M.T., WARBURTON, M.L. AND BUCKLER, E.S. Empirical Comparison of Simple Sequence Repeats and Single Nucleotide Polymorphisms in Assessment of Maize Diversity and Relatedness. **PLoS ONE**, v. 2, 2007. <http://dx.doi.org/10.1371/journal.pone.0001367>

HOWELER, R.; LUTALADIO, N.; THOMAS, G. **Save and Grow: Cassava: A Guide to Sustainable Production Intensification**. Food and Agriculture Organization of the United Nations Rome, Italy, 2013.

HU, W., JI, C., LIANG, Z. et al. Resequencing of 388 cassava accessions identifies valuable loci and selection for variation in heterozygosity. **Genome Biol.**, v. 22, n. 316 2021. <https://doi.org/10.1186/s13059-021-02524-7>

IAE (Institute for Applied Ecology).

TechNote: Distance and Visualization in Population Genetics. Australia: University of Canberra, 2022.

INCAPER (INSTITUTO CAPIXABA DE PESQUISA, ASSISTÊNCIA TÉCNICA E EXTENSÃO RURAL). **Metereologia**. Disponível em: <https://meteorologia.incaper.es.gov.br/>. Acesso em: 10 jan. 2023.

KARIM, K. Y., IFIE, B., DZIDZIENYO, D., et al. Genetic characterization of cassava (*Manihot esculenta* Crantz) genotypes using agro-morphological and single nucleotide polymorphism markers. **Physiology and Molecular Biology of Plants**, v. 26, p. 317-330, 2020. <https://doi.org/10.1007/s12298-019-00740-x>

KUMAR, S.; BANKS, T. W.; CLOUTIER, S. SNP discovery through next-generation sequencing and its applications. **International journal of plant genomics**, v. 2012, 2012. <https://doi.org/10.1155/2012/831460>

KUMAWAT, G.; KANTA KUMAWAT, C.; CHANDRA, K.; et al. Insights into Marker Assisted Selection and Its Applications in Plant Breeding. In: ABDURAKHMONOV, I. Y. **Plant Breeding: Current and Future Views**. 2021. <https://doi.org/10.5772/intechopen.95004>

LORENZI, J. O. Mandioca. 2. ed. Campinas: CATI, 2012. 129p.

MAMMADOV, J.; AGGARWAL, R.; BUYYARAPU, R.; KUMPATLA, S. SNP markers and their impact on plant breeding. **Int. J. Plant Genomics**, v. 2012; n.728398, 2012. <https://doi.org/10.1155/2012/728398>

MANTEL, N. The Detection of Disease Clustering and a Generalized Regression Approach. **Cancer Research**, v. 27, n. 2, p. 209-220, 1967.

MIRANDA, L. A.; SPINOSA, W. A.; DESTRO, T. M.; et al. Sweet cassava cooking time. **Agronomy Science and Biotechnology**, v. 6, p. 1-16, 2020. <https://doi.org/10.33158/ASB.r109.v6.2020>

MOJENA, R. Hierarchical grouping method and stopping rules: an evaluation. **Computer Journal**, v. 20, p. 359-363, 1977. <https://doi.org/10.1093/comjnl/20.4.359>

MONDINI, L., NOORANI, A., & PAGNOTTA, M. A. Assessing plant genetic diversity by molecular tools. **Diversity**, v. 1, n. 1, p. 19-35, 2009. <https://doi.org/10.3390/d1010019>

NTAWURUHUNGA, P.; DIXON, A. G. Quantitative variation and interrelationship between factors influencing cassava yield. **Journal of Applied Biosciences**, v. 26, p. 1594-1602, 2010.

OLIVEIRA, E. J., OLIVEIRA FILHO, O. S., & SANTOS, V. S. Classification of cassava genotypes based on qualitative and quantitative data. **Genetics and Molecular Research**, v. 14, n. 1, p. 906-924, 2015. <https://doi.org/10.4238/2015.february.2.14>

PAZ, R. B. O., DA COSTA, C. H. M., VIEIRA, E. A., et al. Desempenho agrônomo de cultivares de mandioca de mesa em ambiente do cerrado. **In Colloquium Agrariae**, v. 16, n. 3, p. 37-47, 2020. <https://doi.org/10.5747/ca.2020.v16.n3.a370>

- PEPRAH, B. B., PARKES, E.Y., HARRISON, O. A.; et al. Proximate composition, cyanide content, and carotenoid retention after boiling of provitamin A-rich cassava grown in Ghana. **Foods**, v. 9, n. 12, 2020. <https://doi.org/10.3390/foods9121800>
- PIEPHO, H. P., RICHTER, C., & WILLIAMS, E. Nearest neighbour adjustment and linear variance models in plant breeding trials. **Biometrical Journal**, v. 50, n. 2, p. 164-189, 2008. <https://doi.org/10.1002/bimj.200710414>
- PIERRE, N.; WAMALWA, L. N.; MUIRU, W. M.; et al.. Genetic diversity of local and introduced cassava germplasm in Burundi using DArTseq molecular analyses. **Plos one**, v. 17, n. 1, 2022. <https://doi.org/10.1371/journal.pone.0256002>
- PONTE, C.M. de A. **Épocas de colheita de variedades de mandioca**. 2008. Dissertação (Mestrado em Agronomia) – UESB, Vitória da Conquista, BA, 2008.
- POOTAKHAM, W.; SHEARMAN, J. R.; RUANG-AREERATE, P.; et al. Large-scale SNP discovery through RNA sequencing and SNP genotyping by targeted enrichment sequencing in cassava (*Manihot esculenta* Crantz). **Plos One**, v. 9, p. 1-19, 2014. <https://doi.org/10.1371/journal.pone.0116028>
- PUSHPALATHA, R., GANGADHARAN, B. Is Cassava (*Manihot esculenta* Crantz) a Climate “Smart” Crop? A Review in the Context of Bridging Future Food Demand Gap. **Tropical Plant Biol.**, v. 13, p. 201–211, 2020. <https://doi.org/10.1007/s12042-020-09255-2>
- R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing. Vienna, Austria, 2022.
- RABBI, I. Y., KULAKOW, P. A., MANU-ADUENING, J. A. et al. Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (*Manihot esculenta* Crantz). **BMC Genet**, v. 16, n. 115, 2015. <https://doi.org/10.1186/s12863-015-0273-1>
- RAFFA, J. D.; THOMPSON, E. A. Power and Effective Study Size in Heritability Studies. **Stat Biosci**, v. 8, n. 2, p. 264-283, 2016. <https://doi.org/10.1007/s12561-016-9143-2>.
- RESENDE, M. D. V.; SILVA, F. F.; LOPES, P. S.; et al. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa: Universidade Federal de Viçosa/Departamento de Estatística. 2012. 291 p.
- RESENDE, M.D.V. Correções nas expressões do progresso genético com seleção em função da amostragem finita dentro de famílias de populações e implicações no melhoramento florestal. **Boletim Pesquisa Florestal**, Colombo, n. 22/23, p.61-77, 1991.
- ROMERO-SEVERSON, J.; SMITH, J. S. C.; ZIEGLE, J.; HAUSER, J.; et al. Pedigree analysis and haplotype sharing within diverse groups of *Zea mays* L. inbreds. **Theor. Appl. Genet.**, v. 103, p. 567-574, 2001. <https://doi.org/10.1007/PL00002911>.
- SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. **Akaike information criterion statistics**. Reidel Publishing Company, 1986.
- SALES FILHO, J. B. **Caracterização de cultivares de mandioca (*Manihot esculenta* Crantz) pela morfologia e padrões isoenzimáticos**. 1991. Tese (Doutorado em Genética) - Universidade Federal de Viçosa, Viçosa, 1991.
- SAMUELS, D. C.; WANG, J.; YE, F.; et al. Heterozygosity Ratio, a Robust Global Genomic Measure of Autozygosity and Its Association with Height and Disease Risk. **Genetics**, v. 204, n. 3, p. 893-904, 2016. <https://doi.org/10.1534/genetics.116.189936>

- SCHMIDT, P.; HARTUNG, J.; BENNEWITZ, J.; et al. Heritability in plant breeding on a genotype-difference basis. **Genetics**, v. 212, n. 4, p. 991-1008, 2019. <https://doi.org/10.1534/genetics.119.302134>
- SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, v. 6, p. 461-464, 1978.
- SEBBENN, A.M.; SIQUEIRA, A.C.M.F.; KAGEYAMA, P.Y.; et al. Parâmetros genéticos na conservação da cabreúva – *Myroxylon peruiferum* L. F. Allemão. **Scientia Forestalis**, Piracicaba, n.53, p.31-38, 1998.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3, p. 591–611, 1965. <https://doi.org/10.2307/2333709>
- SILVA, R. D. S.; MOURA, E. F.; FARIAS NETO, J. T. D.; et al. Genetic parameters and agronomic evaluation of cassava genotypes. **Pesquisa Agropecuária Brasileira**, v. 51, p. 834-841, 2016. <https://doi.org/10.1590/S0100-204X2016000700006>
- SILVA, R. S.; CUNHA, E.; de FARIAS NETO, J. T.; et al. Selection of morphoagronomic descriptors for the characterization of accessions of cassava of the Eastern Brazilian Amazon. **Genetics and Molecular Research**, v. 16, n. 2, 2017. <http://dx.doi.org/10.4238/gmr16029595>
- SWARUP, S.; CARGILL, E. J.; CROSBY, K.; et al. Genetic diversity is indispensable for plant breeding to improve crops. **Crop Science**, v. 61, n. 2, p. 839-852, 2021. <https://doi.org/10.1002/csc2.20377>
- TEIXEIRA, P. R. G.; VIANA, A. E. S.; CARDOSO, A. D.; et al. Características morfológicas de quatro variedades de mandioca de mesa cultivadas em Vitória da Conquista-BA. **Revista Raízes e Amidos Tropicais**, v. 10, p. 1-8, 2014. <https://doi.org/10.17766/1808-981X.2014v10n1p1-8>
- VIEIRA, E. A.; FIALHO, J. D. F.; CARVALHO, L. J. C. B. Correlação fenotípica entre caracteres agrônômicos em população segregante de mandioca de mesa. **Revista Ceres**, v. 61, p. 523-529, 2014. <https://doi.org/10.1590/0034-737X201461040011>
- VIEIRA, E. A.; FIALHO, J. de F.; SILVA, M. S.; et al. Variabilidade genética do banco de germoplasma de mandioca da Embrapa Cerrados acessada por meio de descritores morfológicos. **Científica: Revista de Ciências Agrárias**, v. 36, n. 1, p. 56-67, 2008. <https://doi.org/10.15361/1984-5529.2008v36n1p56%20-%2067>
- WANG, W.; FENG, B.; XIAO, J. et al. Cassava genome from a wild ancestor to cultivated varieties. **Nat. Commun.**, v. 5, n. 5110, 2014. <https://doi.org/10.1038/ncomms6110>
- YOKOMIZO, G. K. I.; FARIAS NETO, J. T. D. Caracterização fenotípica e genotípica de progênies de pupunheira para palmito. **Pesquisa Agropecuária Brasileira**, v. 38, n. 67-72, 2003. <https://doi.org/10.1590/S0100-204X2003000100009>
- YONIS, B. O.; PINO DEL CARPIO, D.; WOLFE, M., et al. Improving root characterisation for genomic prediction in cassava. **Scientific Reports**, v. 10, n. 1, 2020. <https://doi.org/10.1038/s41598-020-64963-9>

CAPÍTULO II

Número mínimo de SNPs para estudos de diversidade genética em *Manihot esculenta* Crantz

Resumo: A diversidade genética da mandioca tem sido avaliada por diferentes tipos de marcadores moleculares, dentre eles, os marcadores SNPs (Polimorfismos de Nucleotídeo Único), que são genotipados em todo o genoma. Entretanto, ainda não estão definidos quantos marcadores são necessários para os estudos de diversidade e determinar esse valor pode reduzir o custo e o tempo das análises. Neste trabalho, foi estimado o número mínimo de marcadores SNPs para estudos de diversidade genética em uma população de 87 genótipos de mandioca. Foi usada reamostragem, seguida da comparação dos valores de parâmetros genéticos, dos agrupamentos ou das matrizes de distância calculados com sub-conjuntos de diferentes números de marcadores. Também foram analisadas duas populações simuladas (PS1 e PS2), visando melhor comparação entre os métodos. Os resultados mostraram que: (i) o número de SNPs necessários para o cálculo de parâmetros genéticos foi pequeno, sendo possível recuperar os valores com apenas 50 SNPs se considerado um intervalo de confiança de 95%; (ii) a amplitude dos valores dos parâmetros genéticos diminuiu com o aumento do número de SNPs, sendo seguro escolher qualquer amostra aleatória com 1000 ou mais SNPs; (iii) o número de SNPs necessários para a construção do agrupamento foi maior que aquele indicado para o cálculo de parâmetros, demonstrando que o número de marcadores depende do objetivo do estudo; (iii) o número mínimo de SNPs necessários para o agrupamento foi de 3000 na população de mandioca, 13000 na PS1 e todos os SNPs na PS2, mostrando que o número de SNPs varia com a população e, em alguns casos, apenas a genotipagem de todo o genoma consegue descrever com segurança as relações genéticas entre os indivíduos; (iv) os gráficos de correlação entre matrizes de distância *versus* o número de SNPs conseguem apontar se existe ou não um número mínimo de SNPs, apresentando um platô no primeiro caso; (v) o uso da correlação de matrizes para a identificação do número mínimo de marcadores é uma técnica mais fácil que as demais e indica o número de marcadores suficiente para expressar as relações genéticas com bastante precisão quando considerada uma correlação de 0,99. Assim, este trabalho sugere o número mínimo de 3000 SNPs escolhidos aleatoriamente para a população de mandioca e o uso da técnica de correlação de matrizes para a identificação do número mínimo de marcadores, considerando uma correlação de 0,99.

Palavras-chave: mandioca; marcador molecular; parâmetros genéticos; agrupamento; correlação de matrizes.

1 INTRODUÇÃO

A mandioca (*Manihot esculenta* Crantz) é uma cultura de segurança alimentar, ou seja, serve como alimento básico para a alimentação de populações localizadas em muitos países, como Nigéria, Tailândia, Brasil e Indonésia, que são os maiores produtores mundiais (VISSES; SENTELHA; PEREIRA, 2018; FAOSTAT, 2023). Como a mandioca é produzida principalmente por pequenos produtores e a maioria dos seus consumidores vive em países em desenvolvimento, a cultura não recebeu o cultivo intenso e o melhoramento genético realizado nas culturas de cereais, como milho, trigo e arroz (BART; TAYLOR, 2017; AMELEWORK et al., 2021). Desta forma, a variabilidade genética da espécie ainda é muito conservada, com potencial para o melhoramento da cultura, visando aumento produtivo, adaptabilidade e características favoráveis ao consumo (AMELEWORK; BAIRU, 2022).

A diversidade genética da mandioca tem sido avaliada por marcadores morfológicos, bioquímicos e moleculares, como foi revisado por Amelework e Bairu (2022). Dentre os marcadores moleculares atuais, destacam-se os SNPs (Polimorfismos de Nucleotídeo Único) por sua abundância nos genomas, alto nível de reprodutibilidade e formatos de detecção de alto rendimento (MAMMADOV et al., 2012; DAR; MAHAJAN; SHARMA, 2019).

As tecnologias de seqüenciamento de nova geração tem se tornado cada vez mais eficientes, permitindo a genotipagem de milhares de SNPs (CHUNG et al., 2017). Entretanto, sequenciar SNPs de alta densidade em um grande número de indivíduos acaba aumentando os custos e o tempo analítico (MORIN et al. 2004; NAZARENO et al., 2017). O desenvolvimento de painéis com número reduzido de SNPs pode tornar as análises mais acessíveis, o que é especialmente importante para culturas como a mandioca, cujas pesquisas são realizadas principalmente por instituições públicas de países em desenvolvimento (AMELEWOR e BAIRU, 2021), com limitação de recursos. Porém, para a maioria das espécies, ainda é necessário identificar o número e as características ideais desses marcadores para os estudos de diversidade.

A identificação do número ideal de SNPs tem sido realizada através de técnicas de reamostragem de diferentes números de SNPs, seguida da comparação dos valores de parâmetros genéticos (NAZARENO et al., 2017), dos agrupamentos (TURAKULOV; EASTEAL, 2003) ou do nível de diferenciação entre populações (MORIM; MARTIEN; TAYLOR, 2009; NAZARENO et al., 2017) calculados com os sub-conjuntos de marcadores. Outra técnica usada é a correlação entre matrizes de distância, que tem sido aplicada com o objetivo de facilitar a identificação do número mínimo de marcadores necessários para os

estudos de diversidade em marcadores SSR (ZHANG et al., 2002; YOU; ZHANG; WANG, 2004; YANG et al., 2005; WU et al., 2010) e SNPs (WANG et al., 2020). Entretanto, não é relatado o valor de correlação ideal para a identificação do número mínimo de marcadores. A correlação de 0,90 foi usada por You et al. (2004) e Wu et al. (2010), mas os dendrogramas gerados com os números de marcadores identificados com esse valor de correlação não tinham exatamente os mesmos grupos que os agrupamentos originais.

Neste trabalho, foram usadas técnicas de comparação de parâmetros genéticos e de agrupamentos, assim como a correlação entre matrizes de distância, com o objetivo identificar o número mínimo de marcadores necessários para estudos de diversidade genética em mandioca e comparar os métodos de identificação. Também foram analisadas duas populações simuladas para melhor comparação entre os métodos.

2 METODOLOGIA

2.1 CONJUNTOS DE DADOS SNPs

2.1.1 SNPs genotipados em Mandioca

Foram avaliados 87 genótipos de mandioca pertencentes à coleção de trabalho do Programa de Melhoramento de Mandioca do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo. Os genótipos são provenientes de cultivos de pequenos produtores rurais no estado do Espírito Santo, nos municípios de Castelo, Cachoeiro de Itapemirim, Jerônimo Monteiro, Alegre, Mimoso do Sul, Divino de São Lourenço, Conceição do Castelo, Venda Nova do Imigrante, Afonso Cláudio, Domingos Martins, Sooretama, São Mateus, Jaguaré, Nova Venécia e Aracruz.

Os tecidos foliares de todos os genótipos foram coletados e enviados para a Intertek para a extração do DNA. A genotipagem foi realizada via metodologia DArTseq™ na *Diversity Array Technology Pty Ltd.* em Canberra-Austrália, usando uma combinação de sequenciamento de próxima geração HiSeq 2500 (Illumina, EUA) com a tecnologia DArTseq™, conforme descrito por Sansaloni et al. (2020). As sequências genômicas foram alinhadas ao genoma de referência de mandioca v6.1. Os dados genotipados foram imputados a um painel de 388.349 SNPs usando o Beagle 5.0 algoritmo (BROWNING et al. 2018) e um painel de referência de 20.733, principalmente haplótipos de mandioca da África Oriental, derivados de uma combinação de genotipagem por sequenciamento e DArTseq™. Os

genótipos foram então codificados como 0 (homozigoto para o alelo de referência), 1 (heterozigoto) e 2 (homozigoto para o alelo alternativo). Os marcadores SNPs com mais de 10% de dados ausentes e/ou com uma frequência do alelo menos comum (MAF) abaixo de 5% foram descartados, resultando em 17.816 SNPs de alta qualidade.

2.1.2 SNPs simulados

Foram simulados dois conjuntos de dados com 87 genótipos e 19.000 SNPs usando a função *glSim* do pacote *adegenet* no software R. A função permite indicar o número de SNPs estruturados e não estruturados a serem simulados. Para os SNPs não estruturados, todos os indivíduos são extraídos da mesma distribuição binomial. Para os SNPs estruturados, diferentes distribuições binomiais são usadas para dois grupos simulados e as frequências dos alelos derivados nos grupos A e B são construídas para diferir. Uma população foi simulada sem SNPs estruturados (População Simulada 1 - PS1) e a outra com 1.000 SNPs estruturados (População Simulada 2 - PS2). Após a simulação, os conjuntos de dados foram filtrados, excluindo-se os marcadores com MAF inferior a 5%. Após a filtragem, resultaram 17750 SNPs na PS1 e 17991 SNPs na PS2.

2.2 NÚMERO MÍNIMO DE SNPs PARA O CÁLCULO DE PARÂMETROS GENÉTICOS

Foram calculados os valores médios dos parâmetros genéticos Heterozigosidade observada (H_o), Heterozigosidade esperada (H_e), Conteúdo de Informação Polimórfica (PIC) e Índice de Fixação (F_i) para as três populações. Posteriormente, os arquivos de dados contendo os SNPs genotipados em mandioca e os SNPs simulados em PS1 e PS2 foram usados para a obtenção de amostras com 50, 100, 200, 300, 500, 750, 1000, 1500, 2000 e 4000 SNPs dos 87 genótipos, com 100 réplicas de reamostragem. Cada uma das 100 amostras dos diferentes números de SNPs foi usada para calcular os parâmetros genéticos H_o , H_e , PIC e F_i e os resultados foram comparados com as médias dos parâmetros calculadas com todos os SNPs usando boxplots, considerando um intervalo de confiança de 95% representado pelos entalhes nas caixas.

2.3 NÚMERO MÍNIMO DE SNPs PARA ANÁLISES DE AGRUPAMENTOS

Os SNPs genotipados ou simulados para as três populações foram usados para a obtenção de dendrogramas pelo método UPGMA (*Unweighted Pair Group Method with Arithmetic Mean*). Os arquivos de dados com todos os SNPs das três populações foram então usados para a obtenção de cinco amostras com 50, 100, 200, 300, 500, 750, 1000, 1500, 2000,

3000, 4000, 6000, 8000, 10000, 13000, 15000 e 17500 SNPs dos 87 genótipos. As amostras com 50 SNPs foram usadas para obter dendrogramas pelo método UPGMA e os grupos formados foram comparados com aqueles obtidos com o dendrograma obtido com todos os SNPs. Quando os grupos não coincidiam, o número de SNPs era aumentado. O número mínimo de SNPs foi considerado aquele com o qual foi possível obter um dendrograma com os mesmos grupos obtidos com todos os SNPs.

Todos os agrupamentos foram calculados a partir de matrizes de distância euclidiana e o número de grupos foi identificado pelo critério estatístico de Mojena (1977).

2.4 NÚMERO MÍNIMO DE SNPs POR CORRELAÇÃO ENTRE MATRIZES

Os arquivos de dados foram usados para a obtenção de amostras aleatórias com 50, 100, 200, 300, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, 8000, 10000, 13000, 15000 SNPs dos 87 genótipos, com 100 réplicas de reamostragem. Com estas amostras, foram geradas matrizes de distância euclidiana entre os 87 genótipos e estas matrizes foram comparadas com a matriz gerada com todos os SNPs através do teste de correlação de Mantel (MANTEL, 1967).

Para a realização dos cálculos, foi criado um script no software R que permitiu realizar o processo em três passos:

- 1) Cálculo da matriz de distância com todos os SNPs (matriz geral);
- 2) Função que seleciona réplicas de reamostragem dos números de SNPs indicados; calcula as matrizes de distância para todas as réplicas; compara as matrizes com a matriz geral e; resulta na média da correlação entre matrizes das réplicas para cada número de SNPs;
- 3) Plotagem do gráfico do número de SNPs pelo valor da correlação entre matrizes.

2.5 IDENTIFICAÇÃO DE SNPs ESPECÍFICOS PARA OS ESTUDOS DE DIVERSIDADE GENÉTICA

Foram geradas 10.000 amostras com o número mínimo de SNPs indicado pelas análises anteriores para os 87 genótipos de mandioca. A partir dessas amostras, foram calculadas 10.000 matrizes de distância euclidiana e, estas foram correlacionadas com a matriz formada com todos os genótipos. Os valores resultantes foram ordenados em ordem crescente e foram selecionadas as 1000 amostras com os menores valores de correlação. Foram então selecionados os SNPs mais frequentes nestas amostras. Estes SNPs foram estudados em relação à correlação da matriz de distância com a matriz geral e ao dendrograma formado.

3 RESULTADOS

3.1 CARACTERIZAÇÃO DAS POPULAÇÕES

Os valores médios dos parâmetros genéticos, da distância euclidiana e o número de grupos formado nos agrupamentos são apresentados na Tabela 1 para as populações de mandioca, PS1 e PS2.

Tabela 1. Média dos parâmetros genéticos, média da distância euclidiana e número de grupos formados no agrupamento UPGMA de 87 genótipos com base em marcadores SNPs genotipados em uma população de mandioca e simulados para duas populações (PS1 e PS2).

Parâmetro	Mandioca	PS1	PS2
Ho	0,46	0,33	0,34
He	0,35	0,34	0,34
PIC	0,28	0,27	0,27
Fi	-0,34	0,02	0,00
Distância Euclidiana	0,68	0,84	0,82
Número de grupos	17	2	34

Ho = Heterozigotidade observada; He = Heterozigotidade esperada; PIC = Conteúdo de Informação Polimórfica; Fi = Índice de Fixação.

3.2 NÚMERO MÍNIMO DE SNPs PARA O CÁLCULO DE PARÂMETROS GENÉTICOS

Os boxplots referentes aos efeitos do número de SNPs sobre os parâmetros genéticos Heterozigotidade Observada (Ho) e Esperada (He), e Índice de Fixação (Fi) estão apresentados na Figura 1 para a população de mandioca. Observou-se que, mesmo com 50 SNPs, foi possível recuperar os valores dos parâmetros genéticos calculados com todos os SNPs, porém com o aumento do número de SNPs ocorreu uma redução da amplitude e do intervalo de confiança. Isto ocorreu também para o parâmetro PIC (Conteúdo de Informação Polimórfica) e para com nas populações simuladas (dados não mostrados). A pequena amplitude permite que a seleção da amostra com um mesmo número de SNPs seja aleatória, ou seja, independente da amostra selecionada, o valor do parâmetro genético calculado com ela será muito próximo ao valor calculado com todos os SNPs.

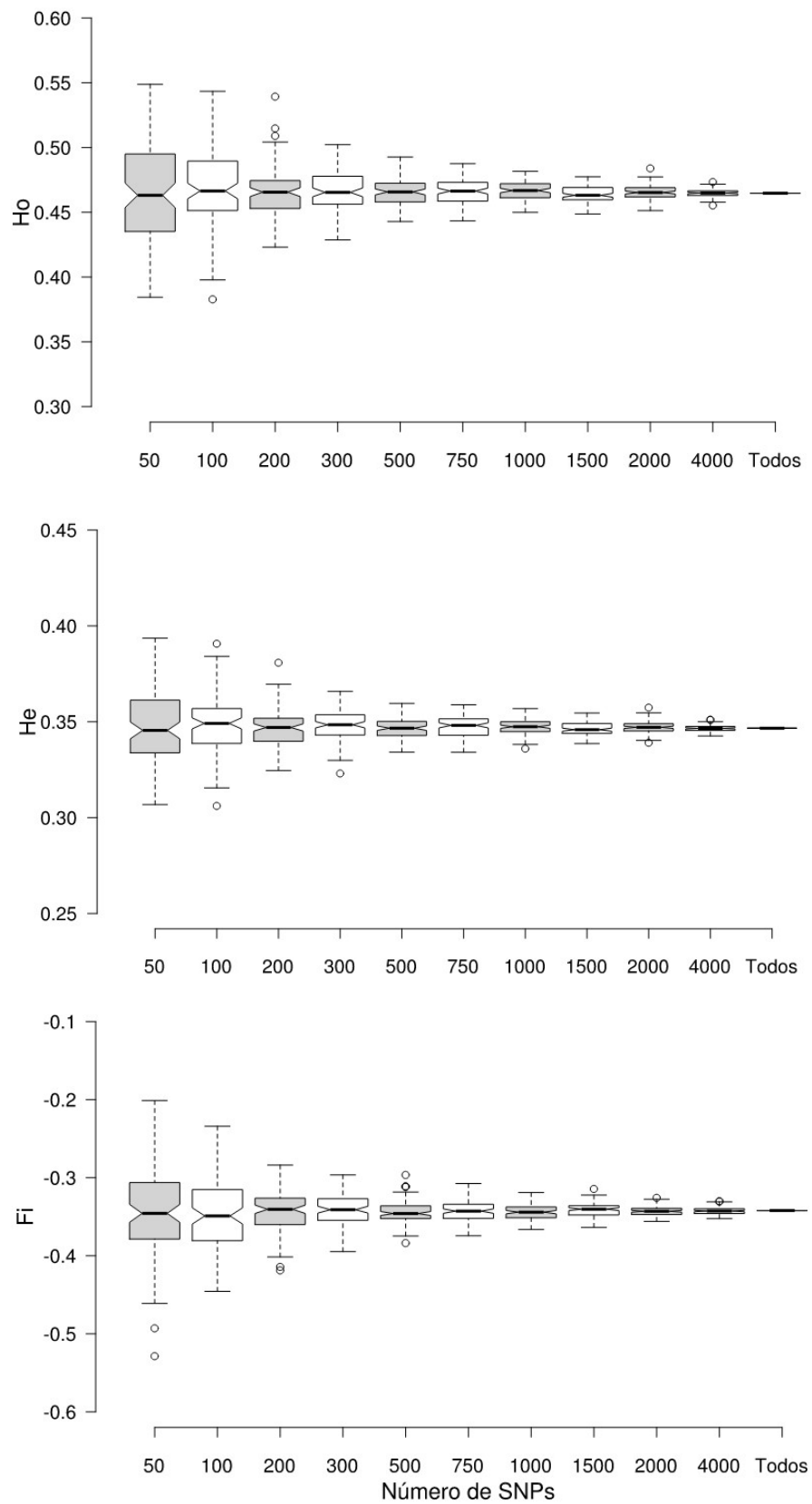


Figura 1. Boxplots dos valores dos parâmetros genéticos Heterozigidade Observada (H_o) e Esperada (H_e) e Índice de Fixação (F_i) calculados com diferentes números de SNPs em uma população de 87 genótipos de mandioca.

3.3 NÚMERO MÍNIMO DE SNPs PARA AGRUPAMENTOS

O agrupamento calculado com todos os SNPs genotipados na população de mandioca está apresentado na Figura 2. Os agrupamentos calculados com todos os SNPs simulados na PS1 e na PS2 estão exibidos nos Apêndices A e B, respectivamente. O número de grupos formados de acordo com o critério de Mojena está apresentado na Tabela 1.

Com os dados de mandioca, foram formados 17 grupos pelo método UPGMA. Com 2000 SNPs algumas amostras permitiram a formação dos mesmos 17 grupos, enquanto outras formaram apenas 16 grupos. Com 3000 SNPs, todas as amostras avaliadas formaram os 17 grupos e esse foi considerado o número mínimo de SNPs.

Para os dados da PS1, foram necessários 13000 SNPs para que todas as amostras formassem os dois grupos com os mesmos genótipos que o agrupamento gerado com todos os SNPs, sendo que, com 12000 SNPs, algumas amostras também permitiram a mesma formação.

No agrupamento dos 87 genótipos da PS2, devido aos valores altos e muito próximos de dissimilaridade entre os genótipos, variando apenas de 80.78 a 83.64, houve a formação de muitos grupos – 35 – e, neste caso, foi observado que só é possível obter um agrupamento semelhante usando todos os SNPs.

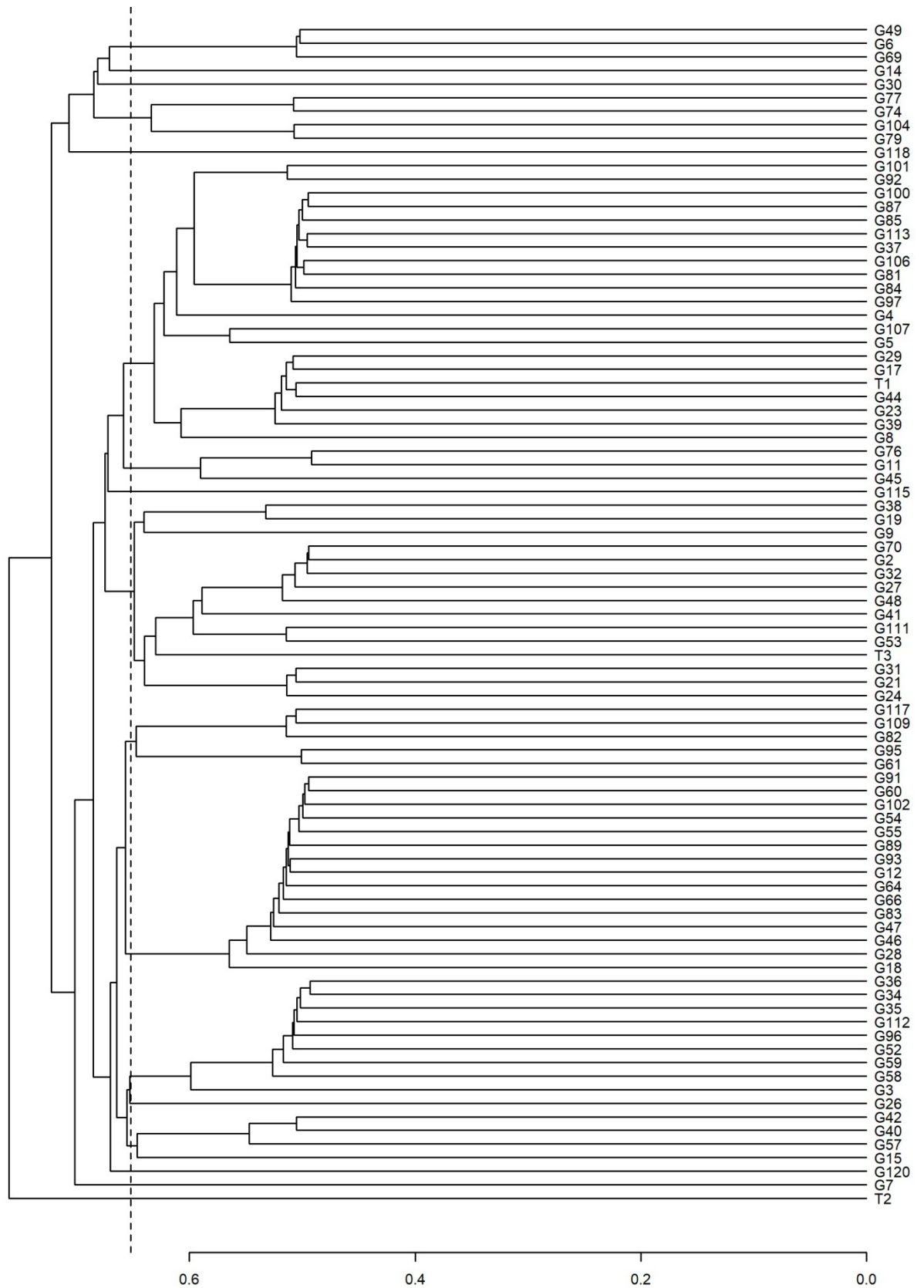


Figura 2. Dendrograma UPGMA da população de mandioca com base em 17816 marcadores SNPs genotipados em 87 genótipos de mandioca. Ponto de corte em 0,65.

3.4 NÚMERO MÍNIMO DE SNPs POR CORRELAÇÃO ENTRE MATRIZES

Diferentes números de SNPs foram amostrados 100 vezes nos arquivos de dados e usados para calcular matrizes de distância euclidiana. Estas sub-matrizes foram então correlacionadas com a matriz gerada com todos os SNPs (matriz geral) e os resultados foram plotados e apresentados na Figura 3.

Foram analisados os números de 50, 100, 200, 300, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, 8000, 10000, 13000 e 15000 SNPs. Entretanto, para os dados da PS2, após observação dos resultados, foram plotados os números de 50, 500, 1000, 2000, 4000, 6000, 8000, 10000, 13000, 15000, 17000 e 17800, para melhor visualização gráfica.

Os gráficos gerados com os dados de mandioca e os dados da PS1 apresentaram uma curva com um rápido aumento da correlação inicialmente, que vai diminuindo e tende a formar um platô. Isso demonstra que, para estes dados, existe um valor a partir do qual, aumentar o número de SNPs não causa elevação representativa na correlação com a matriz geral. Para os dados de mandioca, a curvatura máxima ocorre em 750 SNPs e a curva torna-se praticamente linear a partir de 4000 SNPs. Para os dados da PS1, a curvatura máxima ocorre com 1500 SNPs, entretanto a correlação mantém um aumento bastante significativo até 8000 SNPs.

O gráfico gerado com os dados da PS2 apresentou um comportamento diferente dos demais, mantendo um aumento expressivo nos valores de correlação por toda a curva, o que demonstra que não existe um número menor de SNPs que seja representativo do todo, ou seja, indica que todos os SNPs devem ser avaliados nesse caso.

Considerando os números mínimos de SNPs necessários para agrupamento, de 3000 e 13000 para os dados de mandioca e PS1, respectivamente, os valores médios de correlação seriam de 0,99. Na PS2, este mesmo valor de correlação só é obtido com 17.800 SNPs.

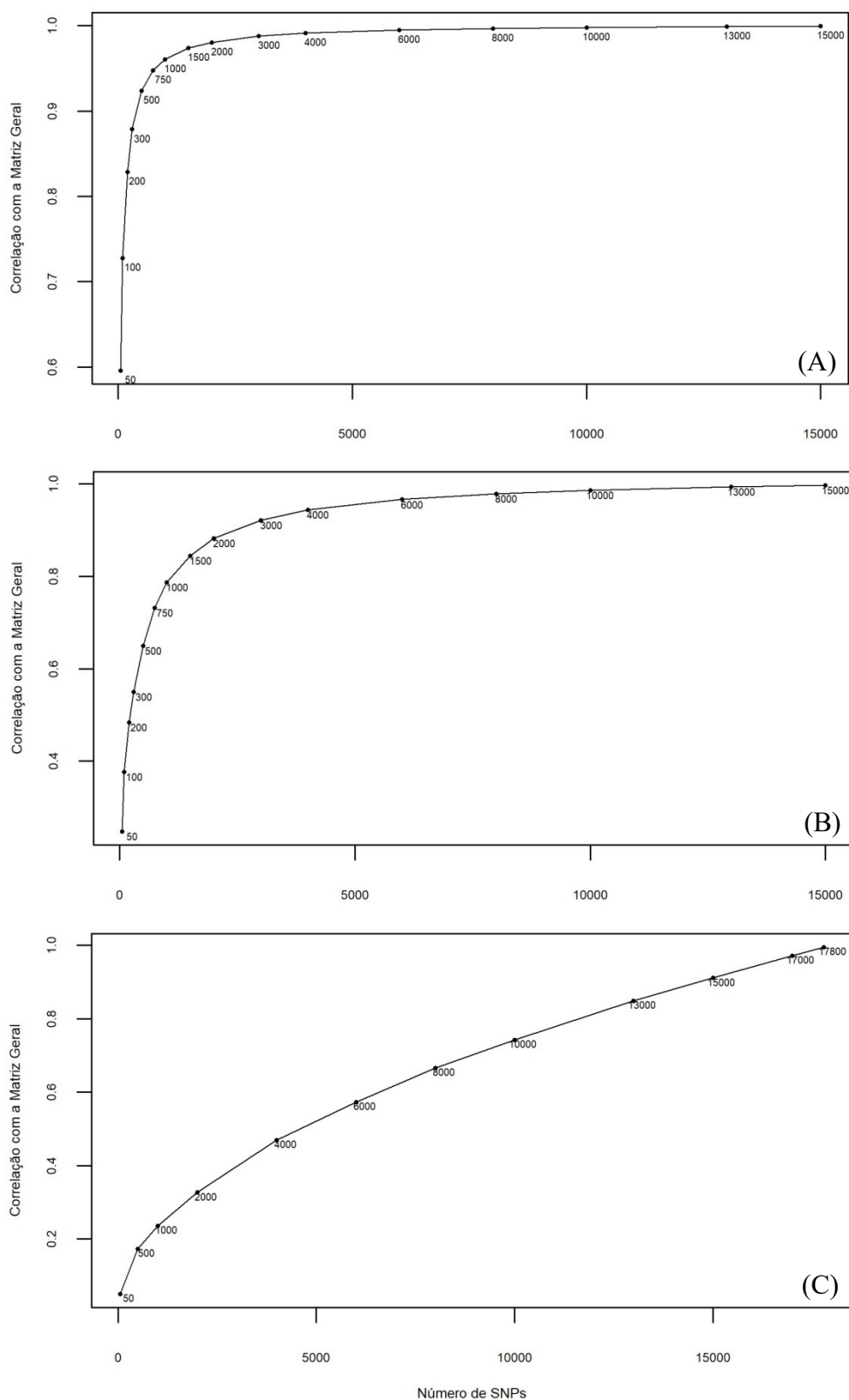


Figura 3. Gráficos de curva revelando a tendência de correlação da matriz de distância com base em diferentes números de SNPs com a matriz gerada com todos os SNPs. (A) População de mandioca; (B) População simulada 1; (C) População simulada 2. As três populações são constituídas de 87 genótipos.

4 DISCUSSÃO

A determinação do número mínimo de marcadores para estudos genéticos é uma oportunidade para a economia de tempo e de recursos dos programas de melhoramento e conservação (MORIM; MARTIEN; TAYLOR, 2009; NAZARENO et al., 2017). Isto é especialmente importante para culturas como a mandioca, cujas pesquisas são realizadas principalmente por instituições públicas de países em desenvolvimento (AMELEWOR e BAIRU, 2021), com limitação de recursos. No caso dos SNPs, após a identificação de um grupo de marcadores informativos e representativos do genoma da espécie, é possível a criação de *arrays* (chips de SNPs) que podem ser usados em substituição ao sequenciamento de todo o genoma, reduzindo os custos das análises.

O número necessário de marcadores, entretanto, ainda não está bem definido para a maioria das espécies e tipos de marcadores e parece ser dependente do objetivo do estudo e da população (SIMKO; EUJAYL; VAN HINTUM, 2012; YUAN et al., 2015; NAZARENO et al., 2017). Neste trabalho, foram estimados os números de SNPs requeridos para análises de quatro parâmetros genéticos e para a construção do agrupamento de uma população de mandioca e de duas populações simuladas.

Para o cálculo de parâmetros genéticos, foi observado que um número baixo de SNPs é necessário, sendo possível recuperar os valores até mesmo com 50 SNPs, com um intervalo de confiança de 95%. Entretanto, considerando uma amostragem aleatória, é mais confiável usar acima de 1000 SNPs, pois, neste caso, qualquer amostra apresentará os valores dos parâmetros genéticos muito próximos aos reais. Valores semelhantes – de 500 e 1000 SNPs – foram necessários para recuperar os valores dos parâmetros genéticos de duas populações de *Amphirrhox longifolia* (NAZARENO et al., 2017). Desta forma, o número mínimo de 1000 SNPs para o cálculo de parâmetros genéticos parece ser suficiente para populações diversas, como as deste trabalho e também aquelas avaliadas pelos autores citados e poderá ser usado como um valor inicial em estudos de outras populações vegetais.

Para as análises de agrupamento, entretanto, é necessário um maior número de marcadores e este é bastante variável de acordo com a população. Para a população de mandioca, 3000 SNPs são suficientes, enquanto para a população PS1 são necessários 13000 e, para a população PS2 são necessários todos os SNPs.

Os marcadores SNPs tem sido usados para a análise da diversidade genética em mandioca através da varredura e identificação dos mesmos em todo o genoma (DE OLIVEIRA et al., 2014; KARIM et al., 2020; PREMPEH et al., 2020). Este estudo

demonstrou que é possível realizar estas análises com um número reduzido de SNPs. Como a população analisada se trata de uma população com alta diversidade genética, representativa do estado do Espírito Santo, SNPs poderão ser selecionados aleatoriamente para a produção de painéis para uso em etapas futuras do programa de melhoramento da mandioca ou mesmo para a genotipagem de novos indivíduos e populações.

A identificação do número mínimo de SNPs pode ser realizada de diversas formas, dependendo do objetivo do estudo. O cálculo dos parâmetros genéticos e de agrupamentos para diversas amostras de diferentes números de SNPs, seguida da comparação entre os resultados, como foi feito neste trabalho e por outros autores (TURAKULOV; EASTEAL, 2003; MORIM; MARTIEN; TAYLOR, 2009; NAZARENO et al., 2017), é bastante laborioso. Usar a correlação de matrizes para identificar esse número mínimo pode ser uma alternativa mais rápida e fácil.

Neste trabalho, observamos que as curvas formadas pelos gráficos do número de SNPs *versus* a correlação com a matriz geral apontam a existência ou não de um número mínimo, apresentando um platô no primeiro caso. Além disso, números de SNPs com correlação igual ou maior que 0,99 parecem ser suficientes para a formação dos agrupamentos. Este valor de correlação corresponde à 3.000, 13.000 e 17.800 SNPs nas populações de mandioca, PS1 e PS2, respectivamente, ou seja, indica os valores necessários para gerar dendrogramas com os mesmos grupos formados com todos os SNPs nas populações de mandioca e PS1 e sinaliza que todos os SNPs devem ser usados na PS2.

O valor de correlação de 0,99 também foi relatado por Zhang et al. (2002) e Rousselle et al. (2015). Zhang et al. (2002), ao avaliar a interferência do número de marcadores SSR no agrupamento de linhagens de trigo, observaram que 450 marcadores possibilitavam a obtenção dos mesmos grupos que todos os 501 marcadores e também relataram uma correlação de 0,99 entre as matrizes de distância geradas por esses números de marcadores. Rousselle et al. (2015), ao avaliar 26.874 SNPs em linhagens de milho, selecionaram 3.072 SNPs que eram suficientes para determinar a cultivar essencialmente derivada e também observaram um valor de correlação de 0,99 entre os dois conjuntos de SNPs.

De acordo com Mantel (1967), uma correlação de 0,90 indica que duas matrizes são altamente consistentes entre si. You et al. (2004) e Wu et al. (2010) consideraram esse valor de correlação para determinar o número mínimo de marcadores SSR, entretanto, relataram que os agrupamentos formados com as sub-matrizes que apresentaram tal valor de correlação não formaram exatamente os mesmos grupos que o agrupamento gerado pela matriz geral, havendo pequenas mudanças de linhas endogâmicas entre grupos. Uma sub-matriz com

correlação de 0,99 provavelmente formaria um dendrograma com os mesmos grupos que a matriz geral.

Em arroz, Yang et al. (2005) perceberam que o dendrograma formado com uma sub-matriz com correlação de 0,90 era completamente diferente daquele formado com a matriz geral. Essa sub-matriz foi calculada com 60 alelos dos 277 totais. Quando os autores usaram 200 alelos, obtiveram uma correlação entre matrizes de 99,5 e um dendrograma praticamente idêntico ao original, com a formação dos mesmos grupos.

Ao estimar o número mínimo de SNPs para determinar a estrutura de uma população com 51 variedades de arroz, Wang et al. (2020) perceberam que 200 SNPs selecionados aleatoriamente eram suficientes para determinar os cinco subgrupos da população indicados pelo total de SNPs de 119.568. Os autores encontraram uma correlação de 0,95 entre a matriz de distância calculada com os 200 SNPs e aquela calculada com todos os SNPs, entretanto não foram relatadas informações em relação ao estudo de diversidade genética e obtenção de dendrogramas. O número de marcadores exigido para a construção de agrupamentos parece ser maior que aquele necessário para a determinação da estrutura populacional, como foi demonstrado por Simko, Eujayl e van Hintum (2012) em beterraba com marcadores SSR, SNPs e SilicoDArTs, e por Yuan et al. (2015) com marcadores SSR em arroz. Desta forma, é provável que, para a formação do agrupamento, um número maior de SNPs fosse necessário no estudo de Wang et al. (2020), o que corresponderia a um valor de correlação também maior, em torno de 0,99.

Assim, considerando os resultados obtidos neste trabalho e nos demais citados, é possível concluir que a correlação entre matrizes é um bom indicativo do número mínimo de marcadores, desde que seja considerado um valor adequado de correlação. Com uma correlação de 0,99 é possível obter amostras de SNPs representativas, suficientes para um estudo de diversidade genética robusto. Esse valor de correlação é baseado no número mínimo de marcadores para análises de agrupamento e, como este número é constantemente maior que aqueles necessários para os cálculos de parâmetros e de estrutura genética, o valor será suficiente em ambos os casos. Além disso, esse valor de correlação parece ser eficiente para marcadores SNPs e SSR e poderá ser estudado para outros marcadores moleculares e também para marcadores morfológicos, quantitativos e qualitativos. Em mandioca, considerando a correlação de 0,99, pode-se dizer que um valor de 3000 SNPs é suficiente para expressar as relações genéticas entre os principais genótipos de mandioca cultivados no estado do Espírito Santo com 99% de certeza, o que reduziria consideravelmente o custo da análise quando comparado com a genotipagem de todo o genoma.

Neste trabalho foi observado que o número de SNPs necessários para os estudos de diversidade genética variam com a população e o tipo de análise. Para estudos populacionais em que as médias são mais importantes que as relações individuais, números muito baixos de SNPs podem ser suficientes. Entretanto, para expressar as relações genéticas entre indivíduos, parece ser necessário um número maior de marcadores, sendo que, para algumas populações, apenas o uso de todos os SNPs genotipados permite representar estas relações de forma suficientemente fiel. Assim, caso o estudo inclua ambas as análises, o número mínimo de SNPs a ser considerado será provavelmente o número necessário para o agrupamento, que é maior. Este número poderá ser mais facilmente calculado com o método de correlação de matrizes, considerando uma correlação de 0,99. Em mandioca, foi possível reduzir fortemente o número de SNPs, de 17.816 para 3.000, e ainda assim calcular as relações genéticas com 99% de representatividade.

5 CONCLUSÕES

O número de SNPs necessários para o cálculo de parâmetros genéticos foi pequeno para todas as populações analisadas. Se considerado um intervalo de confiança de 95%, é possível recuperar os valores com apenas 50 SNPs e se toda a amplitude dos valores for levada em conta, qualquer amostra aleatória com 1000 ou mais SNPs será suficiente para calcular valores muito próximos aos reais.

O número de SNPs necessários para a construção do agrupamento foi maior que aquele indicado para o cálculo de parâmetros e variou de acordo com a população, demonstrando que o número de marcadores depende do objetivo do estudo e da população analisada.

O número mínimo de SNPs necessários para o agrupamento foi de 3000 na população de mandioca e 13000 na PS1. Na PS2 não foi possível estimar um número mínimo para a construção do agrupamento, sendo que apenas com todos os SNPs foi possível agrupar corretamente os genótipos. Ou seja, em algumas populações, apenas a genotipagem de todo o genoma conseguirá descrever com segurança as relações genéticas entre os indivíduos.

O uso da correlação de matrizes para a identificação do número mínimo de marcadores é uma técnica mais fácil que as demais e indica o número de marcadores suficiente para expressar as relações genéticas com bastante precisão quando considerada uma correlação de 0,99.

6 REFERÊNCIAS

- AMELEWORK, A.B.; BAIRU, M.W. Advances in Genetic Analysis and Breeding of Cassava (*Manihot esculenta* Crantz): A Review. **Plants (Basel)**, v. 11, n. 1617, 2022. <https://doi.org/10.3390/plants11121617>
- AMELEWORK, A.B.; BAIRU, M.W.; MAEMA, O.; VENTER, S.L.; LAING, M. Adoption and Promotion of Resilient Crops for Climate Risk Mitigation and Import Substitution: A Case Analysis of Cassava for South African Agriculture. **Front. Sustain. Food Syst.**, v. 5, n. 617783, 2021. <https://doi.org/10.3389/fsufs.2021.617783>
- BART, R.S.; TAYLOR, N.J. New opportunities and challenges to engineer disease resistance in cassava, a staple food of African small-holder farmers. **PLoS Pathog**, v. 13, n. 5, 2017. <https://doi.org/10.1371/journal.ppat.1006287>
- BROWNING, B. L.; ZHOU, Y.; BROWNING, S. R. A one-penny imputed genome from next-generation reference panels. **American Journal of Human Genetics**, v. 103, p. 338–348, 2018.
- CHUNG, Y. S. et al. Genotyping-by-Sequencing: a Promising Tool for Plant Genetics Research and Breeding. **Horticulture, Environment, and Biotechnology**, v. 58, n. 5, 2017. <https://doi.org/10.1007/s13580-017-0297-8>
- DAR, A. A.; MAHAJAN, R.; SHARMA, S. Molecular markers for characterization and conservation of plant genetic resources. **Indian J Agric Sci**, v. 89, n. 11, 2019.
- DE OLIVEIRA, E.J.; FERREIRA, C.F.; SANTOS, V. S; et al. Potential of SNP markers for the characterization of Brazilian cassava germplasm. **Theor Appl Genet**, v. 127, p. 1423–1440, 2014. <https://doi.org/10.1007/s00122-014-2309-8>
- FAOSTAT. (2023). **Food and Agriculture Organization of the United Nations Statistics**. Disponível em: <<http://www.fao.org/faostat/en/#data/QC>>. Acesso em: 12 jan. 2023.
- KARIM, K. Y.; IFIE, B.; DZIDZIENYO, D.; et al. Genetic characterization of cassava (*Manihot esculenta* Crantz) genotypes using agro-morphological and single nucleotide polymorphism markers. **Physiol. Mol. Biol. Plants**, v. 26, p. 317–330, 2020.
- MAMMADOV, J.; AGGARWAL, R.; BUYYARAPU, R.; KUMPATLA, S. SNP markers and their impact on plant breeding. **Int. J. Plant Genomics**, v. 2012; n.728398, 2012. <https://doi.org/10.1155/2012/728398>
- MANTEL, N. The Detection of Disease Clustering and a Generalized Regression Approach. **Cancer Research**, v. 27, n. 2, p. 209-220, 1967.
- MOJENA, R. Hierarchical grouping method and stopping rules: an evaluation. **Computer Journal**, v. 20, p. 359-363, 1977. <https://doi.org/10.1093/comjnl/20.4.359>
- MORIN, P. A.; MARTIEN, K. K.; TAYLOR, B. L. Assessing the statistical power of SNPs for population structure and conservation studies. **Mol. Ecol. Resour.**, v. 9, n. 6, 2009. <https://doi.org/10.1111/j.1755-0998.2008.02392.x>
- NAZARENO, A. G.; BEMMELS, J. B.; DICK, C. W.; LOHMANN, L. G. Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. **Molecular Ecology Resources**, 17, n. 11, 2017. <https://doi.org/10.1111/1755-0998.12654>

- PREMPEH, R. N. A.; MANU-ADUENING, J. A.; QUAIN, M. D. Assessment of genetic diversity among cassava landraces using single nucleotide polymorphic markers. **Afr. J. Biotechnol.**, v. 19, p. 383–391, 2020.
- ROUSSELLE, Y.; JONES, E.; CHARCOSSET, A.; et al. Study on essential derivation in maize: III. Selection and evaluation of a panel of single nucleotide polymorphism loci for use in European and North American germplasm. **Crop Science**, 55, p. 1170 - 1180, 2015. <https://doi.org/10.2135/cropsci2014.09.0627>
- SANSALONI, C.; FRANCO, J.; SANTOS, B.; et al. Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. **Nat. Commun.**, v. 11, n. 4572, 2020. <https://doi.org/10.1038/s41467-020-18404-w>
- SIMKO, I.; EUJAYL, I.; VAN HINTUM, T. J. L. Empirical evaluation of DArT, SNP, and SSR marker-systems for genotyping, clustering, and assigning sugar beet hybrid varieties into populations. **Plant Science**, v. 184, p. 54-62, 2012. <https://doi.org/10.1016/j.plantsci.2011.12.009>
- TURAKULOV, R.; EASTEAL, S. Number of SNPs Loci Needed to Detect Population Structure. **Hum Hered**, v. 55, n. 3, 2003. <https://doi.org/10.1159/000071808>
- VAN HINTUM, T. J. L. Data resolution: a jackknife procedure for determining the consistency of molecular marker datasets. **Theor Appl Genet.**, v. 115, p. 343–349, 2007. <https://doi.org/10.1007/s00122-007-0566-5>
- VISSES, F. D. A.; SENTELHAS, P. C.; PEREIRA, A. B. Yield gap of cassava crop as a measure of food security - an example for the main Brazilian producing regions. **Food Sec.**, v. 10, 1191–1202, 2018. <https://doi.org/10.1007/s12571-018-0831-2>
- WANG, S.; YE, J.; XU, Q.; et al. Determination of minimum number of random SNP for accurate population classification in rice (*Oryza sativa* L.). **Preprint**, 2020. <https://doi.org/10.21203/rs.3.rs-35906/v1>
- WU, C. L.; LI, S. F.; DONG, B. X.; ZHANG, Q. Q.; ZHANG, C. Q. Determination of the Number of SSR Alleles Necessary for the Analysis of Genetic Relationships Between Maize Inbred Lines, **Agricultural Sciences in China**, v. 9, n. 12, 2010. [https://doi.org/10.1016/S1671-2927\(09\)60270-4](https://doi.org/10.1016/S1671-2927(09)60270-4)
- YANG, Q. W.; CHEN, C. C.; ZHANG, W. X.; SHI, J. X.; REN, J. F. Minimum Number of SSR Alleles Needed for Genetic Structure Analysis of *Oryza rufipogon* Populations. **Chinese J. Rice Sci.**, v. 19, n. 4, 2005.
- YOU, G. X.; ZHANG, X. Y.; WANG, L. F. An estimation of the minimum number of SSR loci needed to reveal genetic relationships in wheat varieties: Information from 96 random accessions with maximized genetic diversity. **Molecular Breeding**, v. 14, p. 397–406, 2004. <https://doi.org/10.1007/s11032-004-0285-4>
- YUAN, X. P.; WANG, C. H.; DENG, H. Z.; et al. Minimum of SSR Markers for Analyzing Genetic Variation of *Oryza sativa* L. **Chin. J. Rice Sci.**, v. 29, p. 578-586, 2015.
- ZHANG, X. Y.; LI, C. W.; WANG, L. F.; WANG, H. M.; YOU, G. X.; DONG, Y. S. An estimation of the minimum number of SSR alleles needed to reveal genetic relationships in wheat varieties. I. Information from large-scale planted varieties and cornerstone breeding parents in Chinese wheat improvement and production. **Theor. Appl. Genet.**, v. 106, n. 1, 2002. <https://doi.org/10.1007/s00122-002-1016-z>

ANEXO

ANEXO I - Identificação numérica, identificação do genótipo, nome comum, origem, região e microrregião de 87 genótipos de mandioca (*Manihot esculenta* Crantz) coletados no Estado Espírito Santo.

Id	Gen.	Nome comum	Origem	Região	Microrregião
1	G2	SM	Castelo	SUL	Central Sul
2	G3	Roxinha	Castelo	SUL	Central Sul
3	G4	Pão-do-Chile	Castelo	SUL	Central Sul
4	G5	Amarelinha	Castelo	SUL	Central Sul
5	G6	Pampinha	Castelo	SUL	Central Sul
6	G7	Cacau comum	Castelo	SUL	Central Sul
7	G8	Córrego da Barata	Castelo	SUL	Central Sul
8	G9	CB	Castelo	SUL	Central Sul
9	G11	Cacau amarela	Castelo	SUL	Central Sul
10	G12	AP	Castelo	SUL	Central Sul
11	G14	Manteiga	Castelo	SUL	Central Sul
12	G15	Cachoeiro	Pacotuba/Incaper	SUL	Central Sul
13	G17	Pão-do-Chile	Pacotuba/Incaper	SUL	Central Sul
14	G18	Incaper 1	Pacotuba/Incaper	SUL	Central Sul
15	G19	Incaper 2	Pacotuba/Incaper	SUL	Central Sul
16	G21	Cacau JM	Jerônimo Monteiro	SUL	Caparaó
17	G23	Cacau	Jerônimo Monteiro	SUL	Caparaó
18	G24	5 minutos	Jerônimo Monteiro	SUL	Caparaó
19	G26	Cacau vermelha	Jerônimo Monteiro	SUL	Caparaó
20	G27	Cacau/amarelinha	Jerônimo Monteiro	SUL	Caparaó
21	G28	Vassourinha branca	Divino de São Lourenço	SUL	Caparaó
22	G29	Amarela DSL	Divino de São Lourenço	SUL	Caparaó
23	G30	Cacau DSL	Divino de São Lourenço	SUL	Caparaó
24	G31	01 DSL	Divino de São Lourenço	SUL	Caparaó
25	G32	01 UFES	Área Exp. UFES/ Alegre	SUL	Caparaó
26	G34	5 minutos	Área Exp. UFES/ Alegre	SUL	Caparaó
27	G35	03 UFES	Área Exp. UFES/ Alegre	SUL	Caparaó
28	G36	04 UFES	Área Exp. UFES/ Alegre	SUL	Caparaó
29	G37	05 UFES	Área Exp. UFES/ Alegre	SUL	Caparaó
30	G38	01 RIVE	Rive/ Alegre	SUL	Caparaó
31	G39	02 RIVE	Rive/ Alegre	SUL	Caparaó
32	G40	03 RIVE	Rive/ Alegre	SUL	Caparaó
33	G41	Cacau/casca roxa CC	Conceição do Castelo	SERRANO	Sudoeste Serrana
34	G42	Cacau/folha roxa CC	Conceição do Castelo	SERRANO	Sudoeste Serrana
35	G44	Mandioca rosa	Conceição do Castelo	SERRANO	Sudoeste Serrana
36	G45	Branca 02 CC	Conceição do Castelo	SERRANO	Sudoeste Serrana
37	G46	Amarela CC	Conceição do Castelo	SERRANO	Sudoeste Serrana

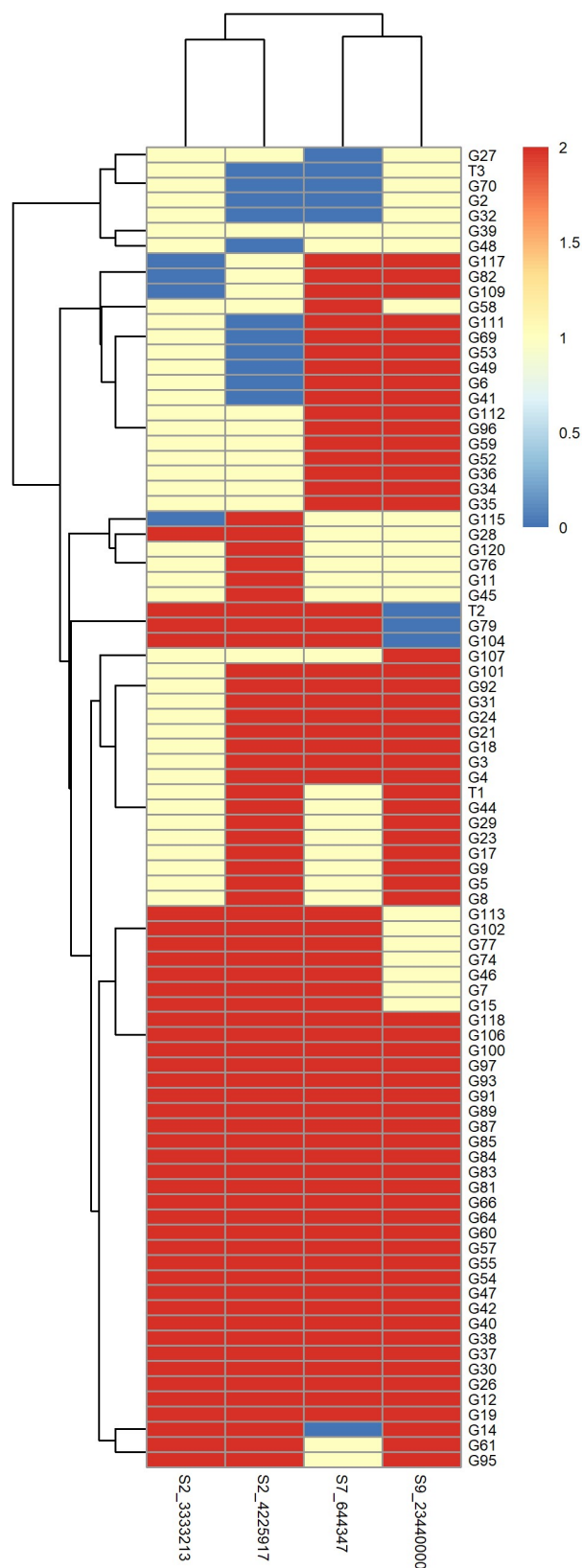
					Continuação
38	G47	Amarela	Conceição do Castelo	SERRANO	Sudoeste Serrana
39	G48	Branca 03 CC	Conceição do Castelo	SERRANO	Sudoeste Serrana
40	G49	Cacau roxa CC	Conceição do Castelo	SERRANO	Sudoeste Serrana
41	G52	Branca 05 CC	Conceição do Castelo	SERRANO	Sudoeste Serrana
42	G53	Cacau 01 CC	Conceição do Castelo	SERRANO	Sudoeste Serrana
43	G54	Mandioca amarela 01	Venda Nova do Imigrante	SERRANO	Sudoeste Serrana
44	G55	Mandioca amarela 02	Venda Nova do Imigrante	SERRANO	Sudoeste Serrana
45	G57	Cacau 01	Venda Nova do Imigrante	SERRANO	Sudoeste Serrana
46	G58	Canoua	Venda Nova do Imigrante	SERRANO	Sudoeste Serrana
47	G59	Cacau amarela 03	Venda Nova do Imigrante	SERRANO	Sudoeste Serrana
48	G60	Cacau cinza	Venda Nova do Imigrante	SERRANO	Sudoeste Serrana
49	G61	Cacau 02	Venda Nova do Imigrante	SERRANO	Sudoeste Serrana
50	G64	Amarelinha AC	Afonso Cláudio	SERRANO	Sudoeste Serrana
51	G66	Roxinha AC	Afonso Cláudio	SERRANO	Sudoeste Serrana
52	G69	01 Domingos Martins	Domingos Martins	SERRANO	Sudoeste Serrana
53	G70	02 DM	Domingos Martins	SERRANO	Sudoeste Serrana
54	G74	06 DM	Domingos Martins	SERRANO	Sudoeste Serrana
55	G76	08 DM	Domingos Martins	SERRANO	Sudoeste Serrana
56	G77	09 DM	Domingos Martins	SERRANO	Sudoeste Serrana
57	G79	11 DM	Domingos Martins	SERRANO	Sudoeste Serrana
58	G81	Ceasa	Sooretama	NORTE	Rio Doce
59	G82	Eucalipto	Sooretama	NORTE	Rio Doce
60	G83	5 minutos	Sooretama	NORTE	Rio Doce
61	G84	Manteiga	Sooretama	NORTE	Rio Doce
62	G85	5 minutos SO	Sooretama	NORTE	Rio Doce
63	G87	Manteiga SO	Sooretama	NORTE	Rio Doce
64	G89	Cacau Nova Aymores	São Mateus	NORTE	Nordeste
65	G91	Amarelinha	São Mateus	NORTE	Nordeste
66	G92	Manteiga Nova Aymores	São Mateus	NORTE	Nordeste
67	G93	Eucalipto Nova Aymores	São Mateus	NORTE	Nordeste
68	G95	Chips Nova Aymores	São Mateus	NORTE	Nordeste
69	G96	Eucalipto SM	São Mateus	NORTE	Nordeste
70	G97	Amarelinha J	Jaguaré	NORTE	Nordeste
71	G100	Cacau J	Jaguaré	NORTE	Nordeste
72	G101	Eucalipto J	Jaguaré	NORTE	Nordeste
73	G102	Aipim Eucalipto	Nova Venécia	NORTE	Nordeste
74	G104	Amarelinha	Nova Venécia	NORTE	Nordeste
75	G106	Manteiga amarela	Nova Venécia	NORTE	Nordeste
76	G107	Aipim manteiga NV	Nova Venécia	NORTE	Nordeste
77	G109	Inca	Aracruz	NORTE	Rio Doce
78	G111	Cacauzinha	Aracruz	NORTE	Rio Doce
79	G112	Racha-terra	Aracruz	NORTE	Rio Doce

					Continuação
80	G113	Rosinha	Aracruz	NORTE	Rio Doce
81	G115	Baiana	Aracruz	NORTE	Rio Doce
82	G117	Alemão Aracruz	Aracruz	NORTE	Rio Doce
83	G118	Cenoura	Aracruz	NORTE	Rio Doce
84	G120	Cacau aldeia	Aracruz	NORTE	Rio Doce
85	T1	Bainha rosa	Mimoso do Sul	SUL	Central Sul
86	T2	Feirinha	Mimoso do Sul	SUL	Central Sul
87	T3	Cacau comum	Mimoso do Sul	SUL	Central Sul

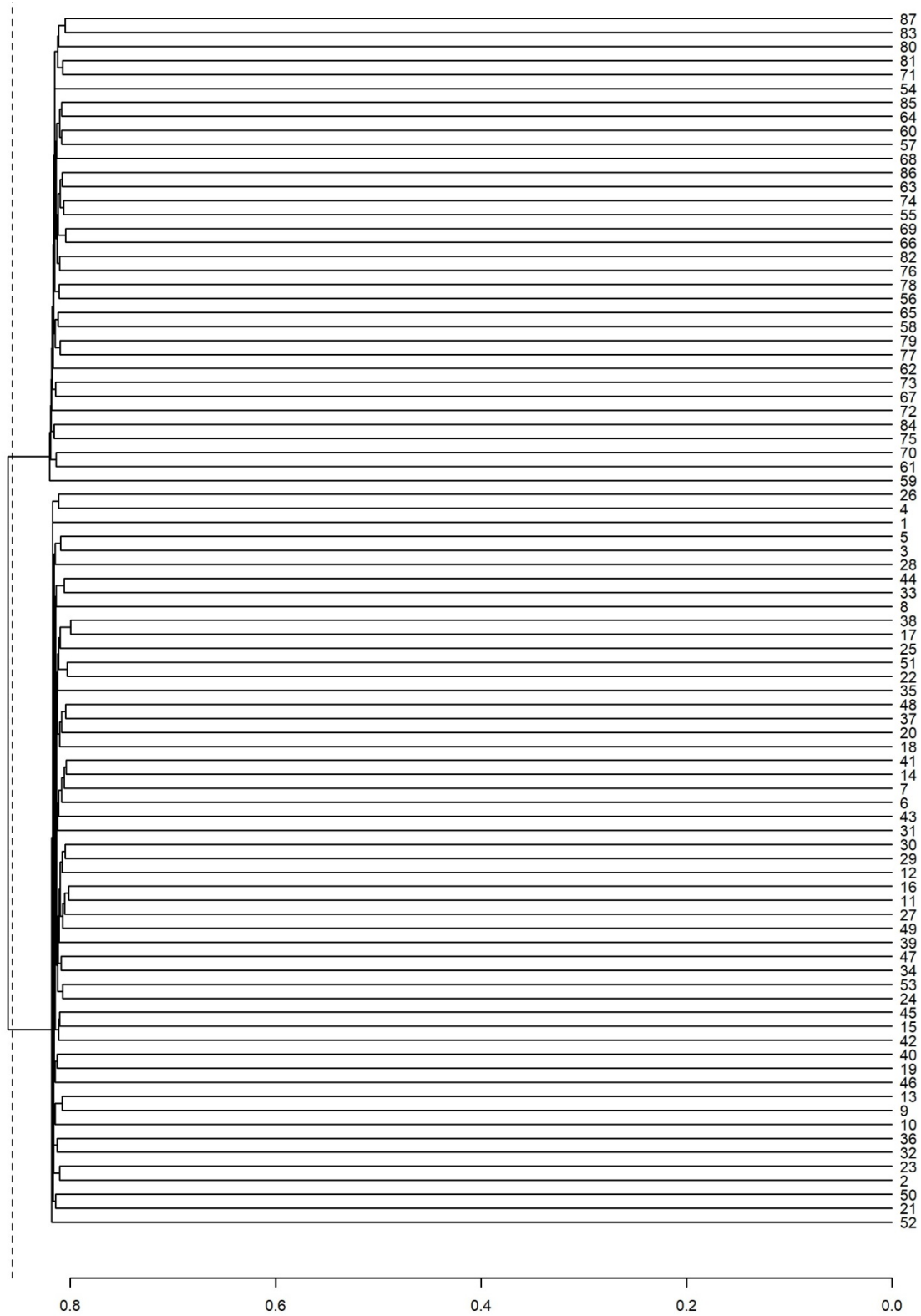
Fonte: Zardo (2022).

APÊNDICES

APÊNDICE A - Heatmap de quatro SNPs cujas matrizes de distância tiveram correlação de 0,48 com a matriz de distância do Peso de Raízes Totais (PESORT) de 87 genótipos de mandioca.



APÊNDICE B - Dendrograma UPGMA da População Simulada 1 (PS1) calculado com 17.750 marcadores SNPs simulados em 87 genótipos. Ponto de corte em 0,86.



APÊNDICE C - Dendrograma UPGMA da População Simulada 2 (PS2) com base em 17991 marcadores SNPs simulados em 87 genótipos. Ponto de corte em 0,82.

