

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

**COLEÇÕES NUCLEARES DE MANDIOCA:  
CONSERVAÇÃO E USO NO MELHORAMENTO DA  
ESPÉCIE**

**Caroline Cardoso dos Santos**

**CRUZ DAS ALMAS - BAHIA**

**Novembro – 2022**

**COLEÇÕES NUCLEARES DE MANDIOCA: CONSERVAÇÃO E USO NO  
MELHORAMENTO DA ESPÉCIE**

Caroline Cardoso dos Santos  
Bacharel em Engenharia Agrônômica, UFRB, 2019.

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

**Orientador:** Eder Jorge de Oliveira  
**Coorientador:** Luciano Rogério Braatz de Andrade

**CRUZ DAS ALMAS - BAHIA  
NOVEMBRO – 2022**

## Ficha Catalográfica

Santos, Caroline Cardoso dos.

Coleções nucleares de mandioca: conservação e uso no melhoramento da espécie/ Caroline Cardoso dos Santos. \_ Cruz das Almas, BA, 2023.

Orientador: Eder Jorge de Oliveira.

Coorientador: Luciano Rogério B. de Andrade.

Dissertação (Mestrado) – Universidade Federal do Recôncavo da Bahia, Centro de Ciências Agrárias, Ambientais e Biológicas.

1.Mandioca – Recursos genéticos vegetais.  
2.Mandioca – Germoplasma vegetal. 3.Variabilidade genética – Análise. Universidade Federal do Recôncavo da Bahia, Centro de Ciências Agrárias, Ambientais e Biológicas. II.Título.

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

**COLEÇÕES NUCLEARES DE MANDIOCA: CONSERVAÇÃO E USO NO  
MELHORAMENTO DA ESPÉCIE**

Comissão Examinadora da Defesa de Dissertação de

Caroline Cardoso dos Santos

Aprovada em 30 de setembro de 2022

Eder Jorge de Oliveira

Universidade Federal do Recôncavo da Bahia – UFRB

Embrapa Mandioca e Fruticultura

Orientador

Cátia Dias do Carmo

Embrapa Mandioca e Fruticultura

Examinador Externo

Cláudia Fortes Ferreira

Embrapa Mandioca e Fruticultura

Examinador Externo

## DEDICATÓRIA

Sou grata a Deus, que me amparou em cada etapa desse trabalho e não me deixou fraquejar e desistir.

À Deus eu dedico.

## AGRADECIMENTOS

Sou grata a Deus por todos os benefícios que tem me proporcionado e por ter me amparado até aqui. Não tenho palavras para descrever a gratidão que tenho ao Senhor por ter me sustentado com sua mão fiel e me ajudado em todos os momentos ruins nessa caminhada.

Sou grata a toda minha família, por todo amor, carinho, cuidado e apoio em todos os momentos que precisei. Em especial à minha mãe Clara e ao meu pai Antônio, por cada palavra sempre dita na hora certa, aos puxões de orelha e incentivos. Aos meus irmãos Cassia Regina e Delfino, que em todos os momentos difíceis nessa caminhada continuaram ao meu lado, apoiando, orando e me acompanhando; não sei o que seria de mim se não fosse o amor e as orações de cada um deles, e aos meus avós, que sempre arrancaram boas risadas e deixaram a minha caminhada mais leve e feliz.

Agradeço ao meu orientador Dr. Eder Jorge, por toda paciência que teve e todos ensinamentos. Saio deste Mestrado com outra visão do mundo acadêmico. Que Deus possa abençoá-lo em tudo na vida e renove suas forças dia após dia. Também não me canso de agradecer ao meu coorientador Luciano, por todo apoio em cada passo e cada etapa, você foi crucial para a finalização deste trabalho.

Não posso deixar de agradecer também a toda equipe Mandioca, em especial a Marcos Souza e o assistente técnico Zé Luiz, que nas etapas de campo, não mediram esforços para me ajudar e me acompanhar em todas as visitas ao BAG, e me instruir em cada etapa de caracterização.

E por fim, agradeço ao programa de Pós-Graduação em Ciências Agrárias, e principalmente aos meus colegas, que sempre estavam dispostos a esclarecer dúvidas e me apoiar em todas as adversidades que enfrentamos nesses anos. Agradeço também à Embrapa Mandioca e Fruticultura pela estrutura física e amparo técnico para desenvolvimento do trabalho.

À Universidade Federal do Recôncavo da Bahia por todo o programa. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Obrigada!

# COLEÇÕES NUCLEARES DE MANDIOCA: CONSERVAÇÃO E USO NO MELHORAMENTO DA ESPÉCIE

## RESUMO GERAL

A mandioca (*Manihot esculenta* Crantz) apresenta uma ampla variabilidade genética conservada em bancos de germoplasma que necessitam de alternativas de uso eficiente. O presente trabalho foi realizado com o objetivo de desenvolver coleções nucleares (capítulo 1) e temáticas (capítulo 2) com base em descritores morfológicos, agronômicos caracterizados em diferentes ensaios entre 2011 e 2021, além de dados moleculares. No Capítulo 1, foram desenvolvidas seis coleções nucleares com diferentes abordagens (EN - *Average entry-to-nearest-entry* e AN - *Average accession-to-nearest-entry*) e diferentes conjuntos de dados (descritores agromorfológicos, moleculares e dados em conjunto). Como não houve uma elevada coincidência na seleção dos genótipos entre as seis coleções nucleares, uma sétima coleção chamada de “coleção consolidada” foi desenvolvida. Todas as coleções mantiveram valores dos parâmetros básicos genéticos bastante próximos a coleção completa. Em relação ao número total de alelos, verificou-se que as coleções CCons e GenEN foram as que conseguiram reter o maior número de alelos nas coleções nucleares (>97%). Na maioria das características não houve uma diferença significativa entre as médias das coleções nucleares comparadas com a coleção completa (>5%). Já a variância foi superior à coleção completa (>50%) nas coleções GPmEN, PhenEN e coleção consolidada para algumas variáveis. As coleções com a melhor distribuição do índice de diversidade de Shannon-Weaver (ISW) foram GPmEN e CCons. Houve boa representação dos genótipos nas diferentes coleções nucleares em comparação com a coleção completa, na análise de componentes principais (PCA). No Capítulo 2, foram desenvolvidas três coleções temáticas para produtividade, resistência a doenças e qualidade de raízes. Os parâmetros genéticos básicos como heterozigosidade observada ( $H_o$ ); diversidade genética dentro da população ( $H_s$ ) e coeficiente de endogamia ( $F_{is}$ ) foram muitos semelhantes, havendo a retenção de >88% dos alelos da coleção completa. Todas as coleções demonstraram um deslocamento de média em todas as características quantitativas nas análises comparativas. A PCA dos dados moleculares e agro-morfológicos demonstraram boa distribuição e elevada dispersão dos genótipos, embora a análise discriminante de componentes principais (ADCP) garantiu maior poder de discriminação dos genótipos em cada coleção, possibilitando observar uma clara separação das três coleções temáticas. Os resultados desse trabalho embasam a escolha da melhor estratégia para formação de coleções nucleares baseada em dados morfo-agronômicos e moleculares e coleções temáticas para atributos produtivos, resistência a doenças e qualidade de raízes. Todas as metodologias adotadas neste trabalho têm grande potencial de uso no melhoramento da cultura.

**Palavras-chave:** *Manihot esculenta* Crantz, germoplasma, coleções temáticas, cultivar.

# CASSAVA CORE COLLECTIONS FOR CONSERVATION AND BREEDING

## GENERAL ABSTRACT

Cassava (*Manihot esculenta* Crantz) exhibits a wide range of genetic variability preserved in germplasm banks, which require efficient utilization strategies. This study aimed to develop core (Chapter 1) and thematic collections (Chapter 2) based on morphological and agronomic descriptors characterized in various field trials between 2011 and 2021, as well as molecular data. In Chapter 1, six core collections were developed using different approaches (EN - Average entry-to-nearest-entry and AN - Average accession-to-nearest-entry) and diverse datasets (agro-morphological descriptors, molecular data, and combined data). As the genotype selection among the six core collections did not demonstrate a high degree of overlap, a seventh collection called the "consolidated collection" was created. All collections exhibited genetic parameters that closely resembled those of the complete collection. Regarding the total number of alleles, the CCons and GenEN collections retained the highest number of alleles (>97%) in the core collections. For most traits, there was no significant difference in means between the core collections and the complete collection (>5%). However, the variance in the GPmEN, PhenEN, and consolidated collections exceeded that of the complete collection (>50%) for certain variables. The GPmEN and CCons collections displayed the best distribution of the Shannon-Weaver diversity index (ISW). The principal component analysis (PCA) showed a good representation of genotypes in the different nuclear collections compared to the complete collection. In Chapter 2, three thematic collections were developed focusing on root yield, disease resistance, and root quality. The basic genetic parameters, such as observed heterozygosity ( $H_o$ ), genetic diversity within the population ( $H_s$ ), and inbreeding coefficient ( $F_{is}$ ), were highly similar among the collections, retaining >88% of the alleles from the complete collection. Comparative analyses revealed a shift in the mean values of all quantitative traits across the thematic collections. Both molecular and agro-morphological data demonstrated a satisfactory distribution and high dispersion of genotypes in the PCA. Furthermore, the discriminant analysis of principal components (DAPC) facilitated better genotype discrimination within each thematic collection, resulting in clear separation among the three collections. The findings of this study provide valuable insights for selecting the best strategies to form core collections based on morpho-agronomic and molecular data, as well as thematic collections focused on yield, disease resistance, and root quality traits. The methodologies employed in this research hold great potential for enhancing cassava breeding programs.

**key words:** *Manihot esculenta* Crantz, germplasm, thematic collection, cultivar.



## LISTA DE ABREVIATURAS

<b>ADCP</b>	Análise discriminante de componentes principais
<b>AN</b>	<i>Average accession-to-nearest-entry</i>
<b>AR</b>	Ângulo de ramificação
<b>AP</b>	Altura de planta
<b>BAG</b>	Banco Ativo de Germoplasma
<b>CAPES</b>	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
<b>Caract</b>	Características
<b>CC_Disease</b>	Coleção temática considerando os principais atributos de resistência a pragas e doenças
<b>CC_Root quality</b>	Coleção temática considerando os principais atributos de qualidade das raízes
<b>CC-Yield</b>	Coleções temáticas considerando os principais atributos de produtividade da mandioca
<b>CCC</b>	Cor do córtex do caule
<b>CCons</b>	Coleção nuclear consolidada
<b>CCR</b>	Cor do córtex das raízes
<b>CE</b>	Comprimento da estípula
<b>CEC</b>	Cor da epiderme do caule
<b>CECC</b>	Cor externa da casca do caule
<b>CER</b>	Cor externa das raízes
<b>CF</b>	Comprimento da filotaxia
<b>CFA</b>	Cor da folha apical
<b>CFD</b>	Cor da folha desenvolvida
<b>CLC</b>	Comprimento do lóbulo central
<b>CMR</b>	Comprimento média das raízes
<b>CN</b>	Cor da nervura
<b>Cont</b>	Continuação
<b>CP</b>	Cor do pecíolo
<b>CP</b>	Comprimento do pecíolo
<b>CPR</b>	Cor da polpa das raízes
<b>CR</b>	Constrições das raízes

<b>CRT</b>	Cor dos ramos terminais
<b>D.AcSev</b>	Severidade causada por ácaro
<b>D.Anth</b>	Severidade causada por antracnose ( <i>Colletotrichum</i> sp.)
<b>D.BacSev</b>	Severidade causada pela bacteriose ( <i>Xanthomonas axonopodis</i> pv. <i>manihotis</i> )
<b>D.FerSev</b>	Severidade da ferrugem ( <i>Uromyces manihotis</i> )
<b>D.FrogSkin</b>	Severidade causada pelo couro de sapo (complexo de vírus e fitoplasmas)
<b>D.MP</b>	Severidade da mancha parda ( <i>Passalora henningsii</i> )
<b>D.Pod</b>	severidade causada pela podridão mole ( <i>Phytophthora</i> sp. <i>Pythium</i> sp. e <i>Phytophythium</i> sp), seca ( <i>Fusarium</i> sp) e negra das raízes ( <i>Neoscytalidium hyalinum</i> e <i>Lasiodiplodia</i> sp.)
<b>D.QF</b>	Severidade da mancha parda ( <i>Passalora viçosae</i> )
<b>DCR</b>	Destaque do córtex das raízes
<b>DMR</b>	Diâmetro médio das raízes
<b>DNA</b>	Ácido desoxirribonucléico
<b>DPR</b>	Destaque da película das raízes
<b>ECR</b>	Espessura da entre casca das raízes
<b>EMBRAPA</b>	Empresa Brasileira de Pesquisa e Agropecuária
<b>EN</b>	<i>Average entry-to-nearest-entry</i>
<b>Fis</b>	Coeficiente de endogamia
<b>FLC</b>	Forma do lóbulo central
<b>FR</b>	Forma das raízes
<b>GBS</b>	<i>Genotyping-By-Sequencing</i>
<b>GenAN</b>	Coleção nuclear formada por dados genotípicas e estratégia de otimização AN ( <i>Average accession-to-nearest-entry</i> )
<b>GenEN</b>	Coleção nuclear formada por dados genotípicos e estratégia de otimização EN ( <i>Averageentry-to-nearest-entry</i> )
<b>GPmAN</b>	Coleção nuclear formada por dados genotípicos e estratégia de otimização EN ( <i>Averageentry-to-nearest-entry</i> )

<b>GPmEN</b>	Coleção formada por dados morfoagronômicos + SNPs e estratégia de otimização EN ( <i>Averageentry-to-nearest-entry</i> )
<b>HCC</b>	Habito de crescimento do caule
<b>HCN</b>	Conteúdo de ácido cianídrico nas raízes
<b>Ho</b>	Heterozigosidade Observada
<b>HR</b>	Habito de ramificação
<b>Hs</b>	Diversidade genética dentro da população
<b>IBS</b>	<i>Identity-by-state</i>
<b>IC</b>	Índice de colheita
<b>ISW</b>	Índices de diversidade de Shannon-Weaver
<b>LLC</b>	Largura do lóbulo central
<b>ME</b>	Margem da estipula
<b>NL</b>	Número de lóbulos
<b>NR</b>	Níveis de ramificação
<b>NR</b>	Número de raízes por planta
<b>P</b>	Pubescência do broto apical
<b>PC</b>	Componentes principais
<b>PCA</b>	Análise de componentes principais
<b>PCF</b>	Proeminência das cicatrizes das folhas
<b>PF</b>	Presença de flores
<b>PhenAN</b>	Coleção nuclear formada por dados fenotípicos e estratégia de otimização AN - <i>Average accession-to-nearest-entry</i>
<b>PhenEN</b>	Coleção nuclear formada por dados fenotípicos e estratégia de otimização AN - <i>Average accession-to-nearest-entry</i>
<b>PI.Por</b>	Porte de planta
<b>PI.RF</b>	Retenção foliar
<b>PI.Vigor1.5M</b>	Vigor das plantas avaliado aos 1,5 meses após o plantio
<b>PL.Vigor12M</b>	Vigor das plantas avaliado aos 12 meses após o plantio
<b>PI.Ty</b>	Tipo de planta
<b>PMS</b>	Produtividade de matéria seca
<b>PP</b>	Posição do pecíolo
<b>PPA</b>	Produtividade da parte aérea
<b>PPR</b>	Presença de pedúnculo das raízes

<b>PR</b>	Posição das raízes
<b>PTR</b>	Produtividade total de raízes
<b>RCL</b>	Relação entre comprimento e largura do lóbulo central
<b>Ro.Cook</b>	Tempo de cozimento das raízes
<b>Ro.Friab</b>	Friabilidade das raízes cozidas
<b>Ro.PRC</b>	Produtividade de raízes comerciais
<b>Ro.PRNC</b>	Produtividade de raízes não comerciais
<b>Ro.TCC</b>	Teor de carotenoides totais nas raízes
<b>S</b>	Sinuosidade das folhas
<b>SNP</b>	<i>Nucleotide Polymorphism Single</i>
<b>StC</b>	Teor de amido
<b>TER</b>	Textura da epiderme das raízes
<b>TMS</b>	Teor de matéria seca
<b>UFRB</b>	Universidade Federal do Recôncavo da Bahia

## LISTA DE TABELAS

### CAPÍTULO 1

**Tabela 1.** Descritores qualitativos avaliados para a formação das coleções nucleares fenotípicas e genotípicas.....52

**Tabela 2.** Resumo das estratégias utilizadas para obtenção das coleções nucleares de mandioca com base em 20.023 marcadores SNPs e 48 descritores agromorfológicos, divididos em qualitativos (33) e quantitativos (15).....56

**Tabela 3.** Estimação dos limites inferiores e superiores das classes fenotípicas a serem utilizadas para a estimação do índice de diversidade de Shannon-Weaver para as características quantitativas.....58

**Tabela 4.** Índice Kappa considerando diferentes metodologias para formação de coleções nucleares.....60

**Tabela 5.** Parâmetros básicos de diversidade genética para as coleções nucleares formadas com base em diferentes abordagens e informações de 20.023 marcadores SNPs.....61

**Tabela 6.** Média e variância para 15 características quantitativas avaliadas em diferentes coleções nucleares de mandioca formadas com base em dados fenotípicos (Phen), genótipos (Gen) e dados fenotípicos e genotípicos agrupadas (GPm) em 1.486 acessos de mandioca.....65

**Tabela 7.** Índices de Shannon-Weaver obtidos com base em 16 descritores quantitativos avaliados em acessos de mandioca para desenvolvimento de seis coleções. ....68

**Tabela 8.** Índices de Shannon-Weaver obtidos com base em 33 descritores qualitativos de folha, caule raiz e flor, avaliados em acessos de mandioca para desenvolvimento de diferentes coleções nucleares.....69

### ANEXOS

**Tabela S1.** Localização geográfica e caracterização de campos dos ensaios do Banco Ativo de germoplasma de Mandioca.....86

**Tabela S2.** Índice Kappa considerando diferentes metodologias para formação de coleções nucleares e coleção consolidada formada a partir dos acessos selecionados por pelo menos duas das abordagens utilizadas.....87

### CAPÍTULO 2

**Tabela 1.** Estimação dos limites inferiores e superiores das classes fenotípicas a serem utilizadas para a estimação do índice de diversidade de Shannon-Weaver

para as características quantitativas.....101

**Tabela 2.** Índice Kappa de coincidência entre acessos selecionados por diferentes coleções temáticas de mandioca.....103

**Tabela 3.** Índices de Shannon-Weaver analisados a partir de descritores quantitativos e qualitativos, considerando três coleções temáticas de mandioca para os principais atributos de produtividade (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality) .....108

**Tabela 4.** Parâmetros básicos de diversidade genética das coleções temáticas formadas com base em diferentes abordagens e informações de 20.023 marcadores SNPs.....109

## **ANEXOS**

**Tabela S1.** Localização geográfica e caracterização de campos dos ensaios do Banco Ativo de germoplasma de Mandioca.....129

**Tabela S2.** Pares de indivíduos com parentesco >0.95 obtidos por IBS (Identity by state) utilizando 20.023 SNPs em 172 indivíduos de mandioca selecionados nas diferentes coleções temáticas (CC\_Root\_quality – Coleção temática para qualidade de raízes; CC\_Yield – Coleção temática para atributos produtivos) .....130

## LISTA DE FIGURAS

### CAPÍTULO 1

**Figura 1.** Número de acessos exclusivos e coincidentes considerando diferentes metodologias para formação de coleções nucleares: Coleções nucleares obtidas exclusivamente com dados fenotípicos e algoritmo de otimização *Average entry-to-nearest-entry distance* (PhenEN) e *Average accession-to-nearest-entry distance* (PhenAN); exclusivamente com dados genotípicos (GenAN e GenEN); e coleções agrupadas com ambos dados fenotípicos e genotípicos (GPmAN e GPmEN).....59

**Figura 2.** Parâmetros genéticos da coleção completa (CCompleta), coleção consolidada (CConsolidada) e coleções nucleares obtidas com base em dados fenotípicos (Phen) com uso do algoritmo *Average entry-to-nearest-entry distance* (AN) e *Average accession-to-nearest-entry distance* (EN) (PhenAN e PhenEN, respectivamente); dados genotípicos (GenAN e GenEN, respectivamente); dados fenotípicos e genotípicos agrupados (GPmAN e GPmEN, respectivamente). **A:** heterozigosidade observada ( $H_o$ ); **B:** heterozigosidade esperada ( $H_s$ ); **C:** coeficiente de endogamia ( $F_{is}$ ).....62

**Figura 3.** Boxplot comparativo entre diferentes descritores quantitativos de mandioca (legenda em material e métodos) considerando diferentes coleções nucleares: Coleção completa (CCompleta), coleções nucleares obtidas com base em dados fenotípicos (Phen) com uso do algoritmo *Average entry-to-nearest-entry distance* (AN) e *Average accession-to-nearest-entry distance* (EM) (PhenAN e PhenEN, respectivamente); dados genotípicos (GenAN e GenEN, respectivamente); dados fenotípicos e genotípicos agrupadas (GPmAN e GPmEN, respectivamente) e coleção consolidada (CConsolidada).....63

**Figura 4.** Barplot comparativo entre diferentes descritores qualitativos de mandioca (legenda em material e métodos) considerando diferentes coleções nucleares: Coleção completa, coleções nucleares obtidas com base em dados fenotípicos (Phen) com uso do algoritmo *Average entry-to-nearest-entry distance* (AN) e *Average accession-to-nearest-entry distance* (EM) (PhenAN e PhenEN, respectivamente); dados genotípicos (GenAN e GenEN, respectivamente), dados fenotípicos e genotípicos agrupadas (GPmAN e GPmEN, respectivamente) e coleção consolidada (CConsolidada).....66

**Figura 5.** Análise de componentes principais (PCA) dos dados fenotípicos de 1486 acessos de mandioca com a dispersão das diferentes coleções nucleares. Coleção completa (CCompleta), coleções nucleares obtidas com base em dados fenotípicos (Phen) com uso do algoritmo *Average entry-to-nearest-entry distance* (AN) e *Average accession-to-nearest-entry distance* (EN) (PhenAN e PhenEN, respectivamente); dados genotípicos (GenAN e GenEN, respectivamente), dados fenotípicos e genotípicos agrupados (GPmAN e GPmEN, respectivamente) e Coleção consolidada que inclui acessos selecionados por pelo menos duas das abordagens utilizadas (Cconsolidada).....71

**Figura 6.** Análise de componentes principais (PCA) com base em 20.023 *single-nucleotide polymorphism* (SNP) de 1486 acessos de mandioca com a dispersão das diferentes coleções nucleares. Coleção completa (CCompleta), coleções nucleares obtidas com base em dados fenotípicos (Phen) com uso do algoritmo *Average entry-to-nearest-entry distance* (AN) e *Average accession-to-nearest-entry distance* (EN) (PhenAN e PhenEN, respectivamente); dados genotípicos (GenAN e GenEN, respectivamente), dados fenotípicos e genotípicos agrupados (GPmAN e GPmEN, respectivamente) e Coleção consolidada que inclui acessos selecionados por pelo menos duas das abordagens utilizadas (Cconsolidada).....72

## ANEXOS

**Figura S1.** Acessos coincidentes considerando a coleção consolidada e as diferentes metodologias para formação de coleções nucleares: Coleções nucleares obtidas exclusivamente com dados fenotípicos e algoritmo de otimização *Average entry-to-nearest-entry distance* (PhenEN) e *Average accession-to-nearest-entry distance* (PhenAN); exclusivamente com dados genotípicos (GenAN e GenEN); e coleções agrupadas com ambos dados fenotípicos e genotípicos (GPmAN e GPmEN).....87

## CAPÍTULO 2

**Figura 1.** Número de acessos coincidentes considerando diferentes coleções temáticas: CC\_Disease – Coleção temática para resistência a pragas e doenças; CC\_Root\_quality – Coleção temática para qualidade de raízes; CC\_Yield – Coleção temática para atributos de produtividade.....102

**Figura 2.** Análise comparativa para diferentes atributos associados aos escores de resistência a pragas e doenças da mandioca, considerando a coleção temática para resistência a pragas e doenças (CC\_Disease) em comparação com a coleção completa. D\_AcSev: ácaros; D\_Anth: antracnose; D\_BacSev: bacteriose; D\_MP: mancha parda; D\_QF: mancha parda; D\_FerSev: ferrugem; D\_FrogSkin: couro de sapo; D\_Pod: podridão radicular.....104

**Figura 3.** Análise comparativa para diferentes atributos associados à qualidade de raízes de mandioca, considerando a coleção temática qualidade de raízes (CC\_Root\_quality) em comparação com a coleção completa. HCN: teor de ácido cianídrico nas raízes; Ro\_TCC: teor de carotenoides totais nas raízes; Ro\_Friab avaliação da friabilidade das raízes cozidas; Ro\_Cook: tempo de cozimento das raízes; CPR: cor da polpa das raízes; CER: cor da casca das raízes; CCR: cor do córtex das raízes.....107

**Figura 4.** Parâmetros genéticos da coleção completa (Completa) e coleções temáticas para resistência a pragas e doenças (CC\_Disease), qualidade de raízes (CC\_Root\_quality) e atributos de produtividade (CC\_Yield). Heterozigosidade observada ( $H_o$ ); heterozigosidade esperada ( $H_s$ ) e coeficiente de endogamia ( $F_{is}$ ).....110

**Figura 5.** Análise de componentes principais (PCA) de 1.810 acessos de mandioca com base em dados moleculares de 20.023 SNP's (A) e dados fenotípicos de 27



variáveis quantitativas (B). Os acessos em cores pertencem a coleções temáticas específicas para atributos produtivos (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality) .....111

**Figura 6.** Análise discriminante de componentes principais (ADCP) de 20.023 SNPs e 27 descritores fenotípicos quantitativas em 1.665 acessos distribuídos diferentes coleções temáticas para atributos produtivos (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality) .....112

## **ANEXOS**

**Figura S1.** Distribuição de valores de compartilhamento de alelos de IBS entre 172 indivíduos que compõem as coleções temáticas de mandioca (CC\_Root\_quality – Coleção temática para qualidade de raízes; CC\_Yield – Coleção temática para atributos produtivos) determinados pela análise de 20.023 SNPs..... 131

## SUMÁRIO

<b>1.0</b>	<b>INTRODUÇÃO GERAL.....</b>	<b>18</b>
<b>2.0</b>	<b>REFERENCIAL TEÓRICO.....</b>	<b>22</b>
2.1	Importância socioeconômica da mandioca.....	22
2.2	Diversidade genética e conservação do germoplasma de mandioca ....	24
2.3	Ações de caracterização do germoplasma de mandioca.....	27
2.4	Coleções nucleares.....	30
2.5	Coleções temáticas.....	33
<b>3.0</b>	<b>REFERÊNCIAS.....</b>	<b>35</b>
	<b>CAPÍTULO 1 – Desenvolvimento de coleções nucleares de mandioca com base em características morfo-agronômicos e marcadores moleculares.....</b>	<b>46</b>
	<b>RESUMO.....</b>	<b>47</b>
	<b>ABSTRACT.....</b>	<b>48</b>
<b>1.0</b>	<b>INTRODUÇÃO.....</b>	<b>49</b>
<b>2.0</b>	<b>MATERIAL E MÉTODOS.....</b>	<b>51</b>
2.1	Material vegetal.....	51
2.2	Descritores morfo-agronômicos.....	51
2.3	Genotipagem dos acessos de mandioca.....	53
2.4	Análise de dados.....	54
2.5	Definição das coleções nucleares.....	55
2.6	Análise de diversidade, comparação e validação dos métodos.....	57
<b>3.0</b>	<b>RESULTADOS.....</b>	<b>58</b>
	Coincidência na seleção dos genótipos para formação das coleções nucleares com base em dados fenotípicos, genotípicos e dados	
3.1	agrupados.....	58
3.2	Formação da coleção consolidada.....	60
	Parâmetros genéticos com base nos dados moleculares da coleção completa, coleções nucleares e coleção consolidada.....	
3.3		61
	Varição nos descritores morfo-agronômicos da coleção completa, coleção consolidada e coleções nucleares.....	
3.4		63
3.5	Análise da diversidade fenotípica das coleções nucleares.....	67
3.6	Validação das coleções nucleares.....	70
<b>4.0</b>	<b>DISCUSSÃO.....</b>	<b>72</b>
	Coincidência de seleção e diversidade dos dados fenotípicos e moleculares .....	
4.1		72
4.2	Maximização da diversidade e representação das coleções nucleares.	76

4.3	Validações das coleções nucleares.....	77
4.4	Aplicação das coleções nucleares na conservação, caracterização e uso dos recursos genéticos da mandioca.....	78
<b>5.0</b>	<b>CONCLUSÃO.....</b>	<b>78</b>
<b>5.0</b>	<b>REFERÊNCIAS.....</b>	<b>79</b>
	<b>ANEXOS.....</b>	<b>86</b>
	<b>CAPÍTULO 2 - Desenvolvimento de coleções nucleares temáticas para atributos produtivos, resistência a doenças e qualidade de amido.....</b>	<b>89</b>
	<b>RESUMO.....</b>	<b>90</b>
	<b>ABSTRACT.....</b>	<b>91</b>
<b>1.0</b>	<b>INTRODUÇÃO.....</b>	<b>92</b>
<b>2.0</b>	<b>MATERIAL E MÉTODOS.....</b>	<b>94</b>
2.1	Dados morfo-agronômicos e avaliação do material vegetal.....	94
2.2	Análise de dados .....	98
2.3	Definição das coleções temáticas.....	99
2.4	Análise de parentesco entre acessos e coincidência entre as coleções...	91
2.5	Parâmetros genéticos das coleções temáticas.....	100
2.6	Diversidade fenotípica e molecular das coleções temáticas de mandioca	101
<b>3.0</b>	<b>RESULTADOS .....</b>	<b>102</b>
3.1	Formação das coleções temáticas.....	102
3.2	Variação fenotípica nas coleções temáticas.....	103
3.3	Análise da diversidade fenotípica das coleções temáticas.....	108
3.4	Análise da diversidade genética de acessos de mandioca baseado em marcadores moleculares.....	109
3.5	Validação das coleções temáticas .....	110
<b>4.0</b>	<b>DISCUSSÃO.....</b>	<b>112</b>
4.1	Formação e diversidade das coleções temáticas de mandioca.....	112
4.2	Representação das coleções temáticas.....	117
4.3	Coleções temáticas para uso prático.....	118
<b>4.0</b>	<b>CONCLUSÕES.....</b>	<b>120</b>
<b>5.0</b>	<b>REFERÊNCIAS.....</b>	<b>121</b>
	<b>ANEXOS.....</b>	<b>129</b>

## 1.0 INTRODUÇÃO GERAL

A mandioca (*Manihot esculenta* Crantz) é uma planta que forma raízes tuberosas ricas em amido, que é o principal produto comercializável, embora outras partes da planta também tenham valor no consumo humano ou animal. Ao longo do processo evolutivo, a mandioca foi submetida a processos de recombinação e ampliação da variabilidade genética consequente da seleção natural, domesticação, facilidade de polinização cruzada e elevada heterozigosidade. Por consequência, há uma infinidade de combinações genotípicas com potencial para adaptar-se às condições edafoclimáticas extremas, que geralmente não são toleráveis para outras culturas, permitindo assim, o seu cultivo em diversas regiões (VENTURINI et al., 2016).

A diversidade da cultura é facilmente observada, nas cores, formato e texturas das raízes; nas formas e cores das folhas e pecíolos; assim como na arquitetura das plantas; na tolerância a diversos fatores bióticos e abióticos (FUKUDA et al., 2005). A avaliação dessa amplitude de características morfológicas são uma importante ferramenta para os programas de melhoramento genético, principalmente nas avaliações preliminares, para compreensão da distribuição genética, sendo crucial para o gerenciamento de coleções de germoplasma (OKOGBENIN et al., 2013).

Em função da importância da cultura para o país, foram criados e vem sendo mantidos bancos ativos de germoplasma, que são unidades conservadoras de material genético de uso imediato ou com potencial de uso futuro (SILVA et al., 2001). No Brasil, existe uma ampla variabilidade genética de mandioca que se encontram conservados em bancos de germoplasma distribuídos por todo país, a maioria desses bancos são mantidos pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e algumas instituições estaduais de pesquisa e desenvolvimento agrícola (NASS, 2007). Existem cerca de 4.000 acessos de mandioca conservados, das quais a maior parte (~2.000 acessos) está armazenada na Embrapa Mandioca e Fruticultura, localizada em Cruz das Almas (Bahia), conservados em condições de campo (*ex situ*) e de laboratório (*in vitro*).

O gerenciamento e utilização dos recursos mantidos em bancos de germoplasma é um desafio, pois a utilização depende diretamente não só da quantidade, mas sim da qualidade dos dados disponíveis (GUO et al., 2014). A

indisponibilidade de informações específicas sobre o germoplasma, como dados de passaporte, caracterização e avaliação, representa uma restrição para utilização dos recursos genéticos da espécie. De acordo com Upadhyaya e Gowda (2009) o número de amostras de germoplasma conservados não corresponde ao número de amostras utilizadas, indicando que os bancos não estão sendo usados em todo o seu potencial. Isso pode ser explicado pelo fato de que as etapas de conservação e caracterização são laboriosas e demandam tempo e recursos financeiros.

Diferentes autores têm sugerido estratégias para melhorar a gestão de grandes coleções e uma das mais utilizadas é a criação de coleções nucleares. Frankel e Brown (1984) defendem que o germoplasma pode ser organizado por meio de coleções com o mínimo de redundâncias possível. Dessa forma, a coleção nuclear é definida como um conjunto reduzido de amostras oriundas de uma coleção de germoplasma que representem a variabilidade genética com o mínimo de repetitividade (BROWN E SPILLANE, 1999). O tamanho reduzido, possibilita que a coleção seja estudada intensivamente e as informações obtidas podem ser usadas para orientar a utilização mais eficiente do germoplasma em programas de melhoramento, bem como no banco de modo geral. Adicionalmente, a coleção nuclear possibilita a completa caracterização e avaliação do germoplasma mais importante, formando uma base completa de informações que promove o uso amplo e eficiente do material genético (CORDEIRO E ABADIE, 2007).

Resumidamente, a coleção nuclear é formada pela quantificação da variabilidade genética total do germoplasma, determinação do tamanho dos subconjuntos principais e seleção dos acessos em cada grupo (PERSEGUINI et al., 2015). Segundo Brown e Spillane (1999), os critérios mais utilizados para implementação das coleções nucleares são a origem geográfica, caracteres morfológicos, grupos taxonômicos, tanto os específicos como os intraespecíficos, regiões ecogeográficas e marcadores genéticos. Diversas metodologias e estratégias auxiliam na formação das coleções nucleares, que podem ser baseadas em caracteres morfo-agronômicos ou dados moleculares.

Independentemente da característica em análise, existem diversas abordagens estatísticas para a formação de coleções nucleares. O programa Core Hunter, por exemplo, tem demonstrado elevada flexibilidade para atender uma variedade de critérios na formação de coleções nucleares com diferentes propósitos, além de suportar otimização de diferentes tipos de variáveis,

maximizando simultaneamente a dissimilaridade genética e a riqueza alélica (BEUKELAER et al., 2018).

Outro ponto importante é aumento contínuo do banco de germoplasma pela introdução de novos acessos derivados de coletas ou mesmo de variedades desenvolvidas pelos programas de melhoramento. Por isso, o conceito de coleções temáticas tem sido introduzido com o propósito de aumentar a eficiência na identificação de diversidade para características específicas e prioritárias nos programas de melhoramento (BOCZKOWSKA et al., 2016). A ideia principal é que grandes coleções de germoplasma, possam contar com várias coleções temáticas de tamanhos menores, com foco em diferentes características de interesse (PESSOA FILHO et al., 2010).

Considerando que as coleções temáticas focam em características com maior potencial de uso agrônomo, imediato ou futuro, é possível proceder à caracterização completa para todos os descritores das espécies, diminuindo substancialmente a redundância dos acessos. Embora o melhoramento genético de mandioca tenha contribuído com o desenvolvimento de variedades com maior resistência a doenças e estresse abiótico, menor toxicidade e maior valor nutricional das raízes, além do aumento do rendimento e qualidade do amido (LIU et al., 2011), o uso do pool genético da *M. esculenta*, não tem sido completamente explorado pela falta de compreensão da diversidade disponível (FERGUSON et al., 2019). Isso tem ocorrido como resultado de fenotipagem ineficientes e pouco confiáveis, e em muitos casos realizadas apenas com uma pequena parte do germoplasma disponível.

Nesse sentido, coleções nucleares e temáticas consistem em um conjunto de acessos selecionados para representar a diversidade genética da coleção completa com o mínimo de repetitividade, garantindo a máxima variação genética, além da rápida avaliação do germoplasma e melhor acesso a coleção completa (CORDEIRO et al., 2000). Portanto, o objetivo geral desse trabalho foi criar coleções nucleares e temáticas de mandioca com base em dados morfológicos, agrônomo e moleculares para desenvolvimento de estratégias de conservação e uso do germoplasma visando desenvolvimento de novas variedades. No primeiro capítulo o objetivo foi: i) desenvolver coleções nucleares de mandioca com base em dados quantitativos, qualitativos e moleculares de forma isolada, ii) avaliar a eficiência de diferentes métodos de seleção de coleções nucleares de mandioca na

retenção de máxima diversidade genética, variância e demais parâmetros genéticos em comparação com a coleção completa e iii) gerar uma coleção nuclear com base em descritores fenotípicos e moleculares de forma conjunta. No Capítulo 2, o objetivo foi desenvolver coleções nucleares temáticas que possam armazenar a máxima diversidade possível para atributos relacionados à produtividade, resistência a doenças e qualidade das raízes, e obter coleções temáticas voltadas para o desenvolvimento de variedades industriais, de mesa e biofortificadas.

## 2.0 REFERENCIAL TEÓRICO

### 2.1 Importância socioeconômica da mandioca

A mandioca é uma planta arbustiva com raízes tuberosas, que pertence à família Euphorbiaceae; um grupo com cerca de 100 espécies do gênero *Manihot* (HOWELER et al., 2013). A espécie é nativa da América do Sul (OLSEN, 2004), mas espalhou-se pelo mundo a partir do século XV, após a colonização da América pelos europeus, tornando-se popular e cultivada em diversas regiões tropicais e subtropicais (CEBALLOS et al., 2010).

Nas últimas décadas, a mandioca foi mundialmente difundida e amplamente cultivada. Tornou-se uma das principais fontes de energia, considerada relevante para segurança alimentar mundial, principalmente para os países subdesenvolvidos e em desenvolvimento (EL-SHARKAWY, 2012). Por isso, a cultura está presente na alimentação básica de cerca de 800 milhões de pessoas, sobretudo na África, América Latina e Ásia (ADEYEMO et. al., 2017). Em muitas regiões, o cultivo da mandioca possui uma relação estreita com a agricultura familiar e de pequena escala, por ser explorada por pequenos produtores em áreas marginais, sobretudo por ser resiliente, capaz de sobreviver em condições de baixa pluviosidade e fertilidade, em que outras culturas não teriam o mesmo desempenho (KARIM et al., 2020).

A mandioca possui elevada tolerância a estresses abióticos, por conseguir proteger as folhas de desidratações severas, e ao mesmo tempo, manter-se fotossinteticamente ativa (EL-SHARKAWY, 2012). Entretanto, por possuir certa tolerância à seca, solos ácidos degradados e oferecer flexibilidade na colheita, em muitas situações, os agricultores não fornecem requisitos mínimos para uma produção competitiva (CEBALLOS et al., 2010).

O maior produtor mundial de mandioca é a Nigéria, com uma produção de 59,5 milhões de toneladas (FAOSTAT, 2020). Segundo a CONAB (2021), a produção brasileira de mandioca tem sofrido reduções anuais, saindo de 21,4 milhões de toneladas em 2014 para 18,49 milhões de toneladas em 2021. Isso tem ocorrido, sobretudo, pela competição com grandes culturas por áreas de plantio, a exemplo da soja e milho, cuja rentabilidade tem aumentado muito nos últimos 10 anos. No Brasil, o consumo da mandioca está bastante associado à população de



baixa renda, e vem passando por um processo de substituição por outros alimentos básicos, a exemplo dos grãos, em populações de maior poder aquisitivo (IRIANTO et al., 2021). Por outro lado, na Indonésia, o consumo de produtos derivados de mandioca tem crescido, seja para uso na alimentação humana, produção de biocombustível ou como ração animal (MAULANA et al., 2020; TAMBUN et al., 2020).

No Brasil, as condições climáticas têm prejudicado o plantio e a colheita da mandioca nos últimos anos, gerando o abandono e a substituição da cultura por cereais (CONAB, 2021), promovendo a redução de 2,5% na área de plantio da mandioca, porém mantendo a produtividade média de 15 t.ha<sup>-1</sup> (IBGE, 2020). Entretanto, existe grande discrepância na produtividade de raízes nas regiões sul, sudeste, norte e nordeste, cujos registros apontam produtividades de 20, 18, 14 e 9 t.ha<sup>-1</sup>, respectivamente (IBGE, 2022).

A mandioca é considerada como um recurso de grande valor para a alimentação nos trópicos, principalmente por apresentar ampla versatilidade quando as possibilidades de uso. A parte aérea e as raízes da mandioca podem ser utilizadas para consumo humano e animal, as raízes possibilitam ainda a agregação de valor sobretudo na forma de farinha e fécula que são matéria prima para fabricação de dezenas de outros produtos (DE SOUZA et al., 2020). As folhas da mandioca têm sido utilizadas na alimentação humana devido às altas concentrações de vitaminas B1, B2 e C, proteínas, carotenoides e minerais (como potássio, magnésio e cálcio), entretanto, tem baixa concentração de zinco, cobre, sódio e ferro (LATIF, MÜLLER, 2015). No entanto, a presença de fatores antinutricionais como oxalato, nitrito e nitrato, que são substâncias mutagênicas, carcinogênicas e teratogênicas, impedem o consumo *in natura* das folhas e raízes pelo homem, pois interferem na absorção de nutrientes no corpo (JESUS et al., 2011).

A espécie possui papel importante na segurança social e econômica, pois permite a colheita das raízes em momentos de maior necessidade nutricional ou demanda, garantindo maior rentabilidade ao agricultor. Recentemente, com o processo de industrialização e modernização da agricultura brasileira, a mandioca tem ganhado bastante importância econômica, devido à extração de amido das raízes, que possui diversas aplicações industriais e culinárias. A agricultura familiar ainda é responsável por cerca de 80% da produção de raiz de mandioca no Brasil

(IBGE, 2017), visando consumo próprio, venda local ou regional, e em muitos casos, a comercialização em agroindústrias. Além disso, o amido e seus derivados são matéria-prima das indústrias, alimentícia, farmacêutica, têxtil e madeireira. Adicionalmente, as sobras, como lascas e pellets, podem ser reaproveitadas na alimentação animal e produção de álcool (DANKWA, PEPRAH, 2019).

## **2.2 Diversidade genética e conservação do germoplasma de mandioca**

A mandioca foi domesticada nas planícies da América do Sul há pelo menos 10.000 anos (OLSEN, 2004). Olsen (2004), utilizou marcadores *single nucleotide polymorphism* (SNP) e microssatélites (SSR) visando rastrear a origem evolutiva e geográfica, concluiu que possivelmente o centro de diversidade da mandioca é a fronteira sudoeste da bacia amazônica, pois todos os marcadores e alelos associados à espécie domesticada também estão presentes na subespécie selvagem *Manihot esculenta* ssp. *flabellifolia*, indicando que possivelmente a mandioca tem um único progenitor selvagem.

A mandioca possui uma ampla variabilidade fenotípica. Zago et al. (2017) reportaram que as características morfológicas são uma importante ferramenta nas avaliações preliminares de germoplasma. Além de auxiliar nos programas de melhoramento genético, são importantes para compreensão da distribuição da diversidade genética; crucial para gerenciamento de coleções de germoplasma (OKOGBENIN et al., 2013). A diversidade da cultura é facilmente observada, por exemplo, nas cores, formato e texturas das raízes; nas formas e cores das folhas ramos, pecíolos; assim como na arquitetura das plantas; na tolerância a diversos fatores bióticos e abióticos.

Com o intuito de padronizar os dados de caracterização do germoplasma de mandioca, listas de descritores foram desenvolvidas por Fukuda et. al. (2010), que englobam características tanto descritivas quanto de valor agrônomo, com sua maioria qualitativa (51) e algumas quantitativas (24). A lista conta com descritores de folha, relacionados desde a cor até a sinuosidade e número de lobo; descritores de caule, relacionados com a cor externa e interno caule, comprimento da filotaxia, habito de crescimento, entre outros; descritores de raiz, relacionados a forma, textura, cores e forma das raízes; descritores de flores frutos e sementes, relacionados ao número de flores, sementes e frutos, ao sexo das flores e presença

ou ausência desses órgãos; e descritores agronômicos, que estão ligados a produtividades dessas plantas, que englobam características como vigor inicial, rendimento das raízes, índice de colheita, conteúdo de ácido cianídrico nas raízes, entre outros.

A conservação da diversidade genética existente da espécie é necessária para atender as demandas da sociedade a curto e longo prazos e garantir uma exploração econômica viável. Nesse intuito, foram criados os bancos de germoplasma (BAG) de mandioca na América Latina e em alguns países do continente africano e asiático (MALIK et al., 2020), com a função de armazenar a maior quantidade da diversidade genética para a espécie.

O centro Internacional de Agricultura Tropical (CIAT) conta com uma coleção de 6.155 acessos de *M. esculenta* de origem asiática, sul-americana, e de parentes silvestres, enquanto que o Instituto Internacional de Agricultura Tropical (IITA) conserva 4.359 acessos de origem africana. Segundo Kawuki et al. (2013), todo material conservado do BAG-mandioca da África está sub-representado nas coleções internacionais. Já no Brasil, existem cerca de 4.000 acessos de mandioca conservados, das quais a maior parte (~2.000 acessos) está armazenada na Embrapa Mandioca e Fruticultura, localizada em Cruz das Almas, Bahia. A grande diversidade genética tem sido reportada por meio de diversos trabalhos, visando a caracterização de acessos com tolerância ao déficit hídrico (SANTOS et al., 2019), tolerância a pragas e doenças (OLIVEIRA et al., 2004; BOARI et al., 2018); propriedades do amido (OLIVEIRA et al., 2018) e caracterização morfo-agronômica. (OLIVEIRA, et al., 2014b), além de estudos relacionados à diversidade genética (ALBUQUERQUE et al., 2018).

Os centros de pesquisas são os grandes responsáveis por coletar, introduzir e caracterizar germoplasma em bancos de conservação. Adicionalmente, os pequenos agricultores e indígenas, também exercem um papel importante na garantia da diversidade agrícola e alimentar, tanto para o valor sociocultural, quanto na conservação da biodiversidade das lavouras (FAO, 2010). A agricultura de pequena escala demonstra grande importância na manutenção da diversidade agrícola, inclusive da mandioca, sendo que comunidades tradicionais ainda mantêm uma considerável diversidade (CARRASCO et al., 2016). Nesse sentido, a conservação *on farm* apresenta essa particularidade, e envolve a conservação de recursos genéticos, principalmente de variedades crioulas, sendo especialmente

cultivadas por pequenos agricultores, em comunidades locais, tradicionais e populações indígenas. Portanto, a estratégia *on farm* é uma estratégia complementar à conservação *ex situ*.

Os principais métodos de conservação para a mandioca são a campo, *in vitro* e em alguns anos a criopreservação vem sendo testada. A metodologia de conservação a campo consiste na renovação do banco de germoplasma de mandioca realizada com manivas do cultivo anterior. A principal desvantagem dessa metodologia está relacionada ao alto custo operacional por necessitar de grande mão de obra empregada em todo ciclo da planta, desde o plantio até o replantio. Além da vulnerabilidade da cultura a diversos fatores, como erosão genética, podendo ocorrer pela não adaptação da espécie às novas condições ambientais, a exposição a germoplasma valiosos a ataques de pragas, doenças e possíveis predadores, intempéries edafoclimáticas, e perda de acessos por falha de identificação decorrente de erro humano ou problemas de cunho técnico-admirativos (SOUZA et al., 2009).

A conservação *in vitro*, por outro lado, consiste na manutenção dos acessos em condições controladas de fotoperíodo, temperatura em meio de cultura que possibilite o crescimento lento, permitindo, assim, a conservação do germoplasma vegetal por um médio período de tempo. Além da vantagem da conservação livre dos riscos existentes no campo, a estratégia *in vitro* viabiliza a manutenção de grande número de acessos em um pequeno espaço físico e o intercâmbio de germoplasma com maior segurança (SILVA et al., 2016). No entanto, tem a desvantagem de exigir uma infraestrutura mais sofisticada e mão de obra especializada, e existe ainda, o grande risco de ocorrência de variação somaclonal, que acontece quando os acessos são conservados *in vitro* e passaram por sucessivos subcultivos, durante longos períodos de tempo (ENGELMANN, 1997).

Para armazenamentos de longo prazo, a criopreservação é uma das alternativas mais viáveis atualmente para a maioria das culturas (ENGELMANN; DUSSERT, 2013). Engelmann e Dussert (2013) afirmaram que em temperaturas muito baixas (-196 °C), todos os processos metabólicos e divisões celulares são interrompidos, assim o material genético pode ser armazenado sem alterações por longo período de tempo. A técnica tem se mostrado muito eficiente para diversas culturas, porém, os estudos de criopreservação em mandioca ainda são inconclusivos para uso generalizado na conservação do germoplasma,

considerando as atuais taxas de regeneração (<50%) (ZHOU et al., 2012) e alguns casos de deformações, como falha no desenvolvimento das raízes de plântulas (PLUMIER et al., 2019).

### **2.3 Ações de caracterização do germoplasma de mandioca**

A caracterização de coleções de germoplasma é uma atividade gerencial fundamental para programas de melhoramento genético, pois envolve a coleta de dados com o objetivo de descrever, identificar e diferenciar componentes de uma mesma espécie e fornecer informações críticas para a preservação e utilização do patrimônio genético recursos. (SALOMÃO, 2010). Em plantas perenes, por exemplo, a caracterização morfo-agronômica executa uma função preponderante para eliminar duplicatas, reduzir custos de manutenção e identificar acessos desejáveis ao melhoramento (BURLE e OLIVEIRA, 2010).

Segundo Burle e Oliveira (2010), as características mais adequadas para uso como descritores são aquelas que são altamente hereditárias, tipicamente controladas por poucos genes, e exibem a mesma expressão em todos os ambientes, ou então exibem pouca interação genótipo-ambiente. Essa caracterização é muito útil para identificar o germoplasma e fornecer uma medida da integridade genética dos acessos que estão sendo conservados.

A caracterização fornece também informações úteis para o manejo do germoplasma em coleções, por meio de descritores que representam caracteres de importância para uso do germoplasma, como, por exemplo, hábito de crescimento, período de floração, frutificação e ciclo. Os descritores são agrupados em listas específicas para cada cultura ou grupo de espécies em questão, e são atribuídos com base em seu estado, ou nas variações aceitas como válidas para aquela característica. (SALOMÃO, 2010).

A caracterização dos acessos de mandioca tem sido realizada com uso de 75 descritores morfológicos e agrônômicos avaliados desde o início de crescimento até o momento da colheita. Atualmente a lista conta com 14 descritores associados às folhas e pecíolos, 15 associados às características de haste, 17 relacionados às raízes, quatro descritores de flores e sementes (FUKUDA et al, 2010). Além de 25 descritores agrônômicos, existem 12 relacionados à produtividade, um relacionado ao conteúdo de ácido cianídrico (BRADBURY et al., 1999), cinco relacionados aos

índices de severidade das doenças e pragas comuns para a cultura da mandioca (ataque de ácaro, queima das folhas, mancha parda, mancha branca, sintomas de antracnose e bacteriose), e sete associados à qualidade de raízes (deterioração fisiológica, cozimento e sabor, e caracterização de grãos de amido) (VENTURINI et al., 2015).

Do ponto de vista econômico, os descritores agronômicos são características importantes que, apesar de sua baixa herdabilidade, contribuem inicialmente para a adaptação e potencial produtivo de cada genótipo (FUKUDA E GUEVARA, 1998). Fukuda e Guevara (1998), confirmaram também que para minimizar os efeitos da interação genótipo x ambiente, muito comum em mandioca, recomenda-se que as observações no campo sejam repetidas entre e dentro de ambientes (anos e locais). Os descritores morfológicos, por outro lado, contêm todas as características que permitem identificar e diferenciar com facilidade do acesso no campo, são características com alta herdabilidade e se expressam em todos os ambientes (FUKUDA E GUEVARA, 1998).

As características morfo-agronômicas são importantes por servir como um primeiro passo para os programas de melhoramento genético, sendo muitas vezes utilizadas em avaliações preliminares. Estudos de caracterização do germoplasma são desenvolvidos com base em dados fenotípicos considerando resistência a doenças, déficit hídrico, qualidade de raízes e atributos produtivos. A caracterização fenotípica do potencial produtivo, resistência a doenças e qualidade das raízes, é um passo fundamental para a seleção de novos clones.

Nos últimos anos, metodologias moleculares que acessam informações diretamente em nível de DNA, têm sido implementadas para permitir inferir com maior precisão aspectos evolutivos, origem genética, identificação de novas variedades, bem como questões de parentesco (GROVER et al., 2014). Essas metodologias envolvem técnicas simples, rápidas e propõem soluções para aumentar a eficácia e confiança nos programas de melhoramento (GARRIDO-CARDENAS et al., 2018). A estratégia molecular, nos tempos atuais, é indispensável em todas as áreas da biologia moderna, pois permite acelerar processos que demandam muito tempo e dinheiro, dispensando, em alguns casos, avaliações fenotípicas nas etapas de campo, e ainda otimizam processos de seleção dos programas de melhoramento (BANGARWA et al., 2021).

As primeiras pesquisas em nível de DNA foram realizadas com marcadores

de polimorfismo de comprimento de fragmentos de restrição (*Restriction Fragment Length Polymorphism* - RFLP), na construção de mapas genéticos (SCHLÖTTERER, 2004; GROVER et al., 2014). Porém, o alto custo e a necessidade de DNA abundante de alta qualidade, limitam esse método, sendo necessário buscar tecnologias alternativas (SCHLÖTTERER, 2004; GROVER et al., 2014; GARRIDO-CARDENAS et al., 2018).

O surgimento de marcadores populares como os microssatélites, *Random Amplified Polymorphic DNA* (RAPD), *Inter Simple Sequence Repeats* (ISSR), e *Amplified Fragment Length Polymorphism* (AFLP), derivados de técnicas baseadas em *Polymerase Chain Reaction* (PCR), geraram uma revolução em diversas avaliações genéticas por eliminar processos exaustivos de genotipagem que demandam muito tempo e recursos financeiros e de infraestrutura (GROVER et al., 2014). A descoberta da PCR possibilitou a geração de mais informações genotípicas com importantes reduções de custo (SCHLÖTTERER, 2004; GROVER et al., 2014). Atualmente, técnicas de detecção de *Single Nucleotide Polymorphism* (SNPs), tem revolucionado a genotipagem pela possibilidade de automação das análises e geração de milhares de dados com custo acessível (SCHLÖTTERER, 2004) possibilitando ampla cobertura genômica (LIU, 2014).

As tecnologias disponíveis e já utilizadas na caracterização molecular em mandioca são as isoenzimas (LESTARI et al., 2017); RAPD (SILVIA et al., 2015); RFLP (BORAH et al., 2012); AFLP (SANTAYANA et al., 2014); TRAP (*Target Region Amplification Polymorphism*) (CARMO et al. 2015), micro (OLIVEIRA et al. 2012) e minissatélites; CAPS (*Cleaved Amplified Polymorphic Sequence*) (VIDYA et al., 2019); SSCP (*Single-Strand Conformation Polymorphism*); ISSR (AFONSO et al., 2019; ASHA et al., 2019); marcadores baseados em retrotransposons (KUHN et al., 2016), e SNPs (*Single Nucleotide Polymorphism*) (OLIVEIRA et al., 2014a; ADU et al., 2020,2021).

Várias técnicas de obtenção de marcadores moleculares foram desenvolvidas e exploradas em todo o mundo (GROVER et al., 2014). Em mandioca, os marcadores moleculares têm sido utilizados na caracterização de germoplasma e estudos de diversidade (ADU et al., 2021, ALBUQUERQUE et. al, 2018); na identificação de duplicatas no germoplasma (ALBUQUERQUE et. al, 2019); na seleção e resistência a patógenos (TEIXEIRA et al., 2021a; TEIXEIRA et al., 2021); na seleção assistida por marcadores (IGE et al., 2021, CARMO et al.

2015); na identificação de genes responsáveis por características de interesse (YOU et al., 2021); *fingerprinting* molecular (OCAMPO et al., 2021); em estudos de associação com propriedades do amido (MA'ARUF et al., 2020, SANTOS et al. 2022) e amido *waxy* (CARMO et al. 2020); teor de carotenoides e de compostos cianogênicos (CHAIAREEKITWAT et al., 2022, CARVALHO et al. 2022).

A incorporação dos avanços recentes de genotipagem com as mais diversas ferramentas moleculares nos estudos de conservação e uso dos recursos genéticos de mandioca para benefício dos agricultores precisa ser implementada de forma efetiva. É preciso integrar todos os tipos de caracterizações e avaliações disponíveis para se estudar e gerenciar de forma mais eficaz a diversidade genética disponível (ASARE et al., 2011).

O gerenciamento e utilização da diversidade mantida em bancos de germoplasma é um desafio, pois a utilização desses recursos depende da quantidade e qualidade dos dados disponíveis (GUO et al., 2014). Etapas de conservação e caracterização são laboriosas, demandam tempo e recursos financeiros para serem executadas. Nesse sentido, o uso de coleções nucleares é uma alternativa para a redução dos custos despendidos nas etapas de conservação e caracterização.

## **2.4 Coleções nucleares**

O termo coleção nuclear é utilizado para designar um conjunto de acessos menor do que o número total que compõe um determinado germoplasma, e que contém a máxima diversidade genética, com a menor redundância possível (FRANKEL, 1984). A obtenção das coleções nucleares pode ser realizada por meio de análises de características agronômicas, morfológicas, bioquímicas e/ou moleculares. Geralmente, as coleções nucleares são bem menores que a coleção principal e em muitos casos não ultrapassam 10% do número total de acessos (BROWN, 1999). Entretanto, na prática, as proporções das coleções são de 5% a 20%, representando até 90% da diversidade de toda a coleção.

Van Hintum et al. (2000), recomendaram a adoção de um exercício de amostragem ordenada em cinco etapas: i) identificação e determinação do material amostrado; ii) definição do tamanho da coleção nuclear; iii) divisão das coleções bases em grupos para se obter a máxima representatividade da diversidade; iv)



definição do número de amostras por grupo visando a determinação da proporção com que cada estrato será representado; v) seleção das amostras nos grupos que serão incluídos na coleção nuclear. Resumidamente, a coleção nuclear principal é formada pela quantificação da variabilidade genética total do germoplasma, determinação do tamanho dos subconjuntos principais e seleção dos acessos em cada grupo (PERSEGUINI et al., 2015). Segundo Brown e Spillane (1999), os critérios mais utilizados para implementação das coleções nucleares são a origem geográfica, caracteres morfológicos, grupos taxonômicos, tanto os específicos como os intraespecíficos, regiões ecogeográficas e marcadores genéticos.

Diversas metodologias e estratégias auxiliam na formação das coleções nucleares, que podem ser baseadas em caracteres morfo-agronômicos ou dados moleculares. Independentemente da característica em análise, existem diversas abordagens estatísticas para a formação de coleções nucleares. Os escores oriundos da *Principal Component Analysis* (PCA) têm sido utilizados para eliminar a colinearidade entre as características e ainda permitir a seleção dos indivíduos com base em sua contribuição cumulativa relativa (NOIROT et al., 1996).

O método de amostragem estratificado, agrupa os acessos de acordo com os valores das matrizes de distância com base em informações moleculares e/ou fenotípicas (BEUKELAER et al., 2018). Ao contrário de outros métodos, a estratificação possui a vantagem de utilizar todas as variáveis simultaneamente, e ainda vincular distância genética ao conceito de diversidade genética (ODONG et al., 2013). Por outro lado, o método *Maximum Length SubTree algorithm* (MLST) implementado no software DARwin, possui várias vantagens quando comparadas com outras abordagens, por utilizar todos os marcadores/características de forma simultânea por meio de uma medida de dissimilaridade ao invés de subconjuntos pré-selecionados, considerando a estrutura populacional (NDJIONDJOP et al., 2017). A priori não requer subgrupos pré-definidos e nem tamanho de amostra fixa, fornecendo assim, indicadores de suporte estatístico detalhados para que os usuários escolham um tamanho de amostra ideal das coleções (NDJIONDJOP et al., 2017).

Os métodos de maximização, como a estratégia M, selecionam combinações específicas de acessos de forma a maximizar a retenção da máxima diversidade. O software MSTRAT utiliza a estratégia M com um algoritmo simples, que utiliza o critério de riqueza alélica e a maior soma de quadrados das

características ativas para selecionar subconjuntos para as coleções nucleares (GOUESNARD et al., 2001). Mais recentemente uma abordagem heurística combinada com a estratégia de maximização foi implementada no software PowerCore (KIM et al., 2007). Esse software identifica coleções nucleares com base na normalização de variáveis contínuas, levando em consideração todas as classes com baixa redundância alélica, garantindo uma lista de genótipos da coleção com elevada reprodutibilidade. O PowerCore é aplicável a diversos tipos de dados genômicos, incluindo SNPs, além de ser uma abordagem que simplifica os processos, reduz o número de entradas e mantém a diversidade categórica (KIM et al., 2007).

A Embrapa Mandioca e Fruticultura já desenvolveu coleções nucleares com base em 798 alelos identificados por 402 marcadores SNPs (OLIVEIRA et al., 2014a). Nesse estudo foi comparada a representatividade de coleções nucleares com diferentes porcentagens de acessos, sendo que uma baixa intensidade de amostragem foi suficiente para representar a diversidade genética de toda a coleção. Foram utilizados diferentes algoritmos baseados na estratégia M, selecionando combinações específicas de acessos, e maximizando a retenção da máxima diversidade. As estratégias empregadas foram padrão M, usando o software MSTRAT e a estratégia M com busca heurística realizada pelo software PowerCore v1.0, e a estrutura genética foi inferida com base nas relações genéticas entre os acessos, avaliadas por análise de componentes principais (OLIVEIRA et al., 2014a). Contudo, essas coleções não contaram com um grande número de descritores e marcadores moleculares como esse estudo propõe, além do mais, o trabalho desenvolvido por Oliveira et al. (2014), teve como principal objetivo testar diferentes porcentagem de tamanho da coleção completa, em termos de representatividade.

Por outro lado, o Core Hunter, como uma ferramenta central e versátil de seleção de subconjuntos, superou as estratégias anteriores por meio de algoritmos de busca local que geram conjuntos com base em métricas como distância e riqueza alélica (BEUKELAER et al., 2018). Ao contrário de outros métodos, o Core Hunter pode construir conjuntos para aplicações específicas e combinar vários objetivos para aproximar diferentes perspectivas, como, por exemplo, maximizar a diversidade genética e a riqueza alélica (Beukelaer et al., 2018). Beukelaer et al. (2018), afirmaram também que a ferramenta Core Hunter pode amostrar núcleos

de tamanho fixo com base em dados de marcadores moleculares, características fenotípicas, matriz de distância pré-computada ou uma combinação destes.

## 2.5 Coleções temáticas

No cenário global atual, com a crescente demanda de ações para o desenvolvimento de novas variedades e a constante redução de recursos financeiros para atividades essenciais como manutenção efetiva dos acessos, preservação da identidade genética, garantia de produção de material livre de patógenos (um dos principais problemas da prorrogação clonal em mandioca) fazem com que a identificação dos genótipos promissores no germoplasma e a posterior concentração dos esforços na sua conservação e caracterização sejam fundamentais (BOUKAR et al., 2013). Moura et al. (2013) relataram que a caracterização e organização dos caracteres para a construção de coleções nucleares é uma abordagem central no manejo dos recursos genéticos de *M. esculenta*, pois reduz o tamanho e facilita o gerenciamento e manuseio do banco de germoplasma. Porém, coleções nucleares sem foco em características ou objetivos específicos para uso no melhoramento genético e ainda com elevado número de genótipos, não são atrativas para uso pelos melhoristas da espécie. Portanto, faz-se necessário a formação de coleções temáticas para servir como material base para escolha de parentais que possam ser utilizados para melhoramento populacional e geração de novas cultivares.

Considerando que as coleções temáticas focam em características com maior potencial de uso agrônomo, imediato ou futuro, é possível proceder à caracterização completa para todos os descritores das espécies, diminuindo substancialmente a redundância dos acessos. Para soja, Guo et al. (2014) propuseram um conceito de coleção nuclear aplicada, que segue o mesmo princípio de coleções nucleares temáticas, formando subconjuntos significativamente representativos para características específicas de interesse, a exemplo de tolerância ao frio, tolerância à seca, resistência ao nematoide de cisto da soja, resistência ao vírus do mosaico da soja, alto teor de proteína e alto teor de gordura, resolvendo a problemática de redução do tamanho amostral e ao mesmo tempo, conservando a diversidade genética.

Pessoa Filho et al. (2010), utilizaram uma metodologia capaz de

potencializar a riqueza alélica nas coleções temáticas com tamanho reduzido. Para isso, foi realizada a seleção prévia de acessos diversos com característica de interesse, como a tolerância à seca e, em seguida, a coleção temática foi obtida levando-se em consideração as distâncias genéticas com base em marcadores moleculares. Para características complexas, diferentes coleções centrais temáticas com um número limitado de acessos podem ser definidas a partir do germoplasma da espécie em estudo, porém com tamanho mais manejável pelos programas de melhoramento (PESSOA FILHO et al., 2010). Este subconjunto temático pode reduzir drasticamente o número de entradas a serem avaliadas e ainda fornecer uma coleção funcional, que pode ser amplamente examinada para todas as características de interesse e, economicamente, importantes, sendo úteis para programas de melhoramento (UPADHYAYA, ORITZ, 2009).

Embora o melhoramento genético de mandioca tenha contribuído com o desenvolvimento de variedades com maior resistência a doenças e estresse abiótico, menor toxicidade e maior valor nutricional das raízes, além do aumento do rendimento e qualidade do amido (LIU et al., 2011), o uso do pool genético da *M. esculenta*, não tem sido completamente explorado pela falta de compreensão da diversidade disponível (FERGUSON et al., 2019). Isso tem ocorrido como resultado de fenotipagem ineficientes e pouco confiáveis, e em muitos casos realizadas apenas com uma pequena parte do germoplasma disponível. Por outro lado, o conceito de coleções temáticas proposto por Upadhyaya et al. (2009a), que indica a seleção de no máximo 10% dos acessos de uma coleção nuclear comum (~ 1% da coleção total) pode ser utilizado para se criar coleções com propósitos específicos, a exemplo de produtividade e qualidade das raízes de mandioca e resistências a doenças.

Pelo menor número de acessos, as coleções temáticas são, portanto, mais manejáveis e por isso permitem extensivas caracterizações para diversas características, incluindo avaliações agronômicas em diferentes anos e locais de cultivo. Isso traz novas perspectivas para obtenção de fenótipos mais confiáveis para diversos estudos genômicos mais aprofundados, como é o caso das abordagens GWAS e de seleção genômica.

### 3.0 REFERÊNCIAS

ADEYEMO, O. S.; CHAVARRIAGA, P.; TOHME J.; FREGENE M.; DAVIS S. J.; SETTER T. L. Overexpression of *Arabidopsis* Flowering Locus T (FT) gene improves floral development in cassava (*Manihot esculenta*, Crantz). **PLoS One**, v. 12, e0181460.2017.

ADU, B. G.; AKROMAH, R.; AMOAH, S.; NYADANU, D.; YEBOAH, A.; ABOAGYE, L. M.; OWUSU, E. G. High-density DArT-based SilicoDArT and SNP markers for genetic diversity and population structure studies in cassava (*Manihot esculenta* Crantz). **PloS one**, v. 16, n. 7, p. e0255290, 2021.

ADU, B. G.; YEBOAH, A.; AKROMAH, R.; BOBOBEE, E.; AMOAH, S.; KENA, A. W.; AMOAH, R. A. Whole genome SNPs and phenotypic characterization of cassava (*Manihot esculenta* Crantz) germplasm in the semi-deciduous forest ecology of Ghana. **Ecological Genetics and Genomics**, v. 17, p. 100068, 2020.

AFONSO, S. D. J.; MOREIRA, R. F. C.; DA SILVA LEDO, C. A.; FERREIRA, C. A. F.; DA SILVA SANTOS, V.; MUONDO, P. A. O. Genetic structure of cassava populations (*Manihot esculenta* Crantz) from Angola assessed through (ISSR) markers. **African Journal of Biotechnology**, v. 18, n. 7, p. 144-154, 2019.

ALBUQUERQUE, H. Y. G.; CARMO, C. D. D.; BRITO, A. C.; OLIVEIRA, E. J. D. Genetic diversity of *Manihot esculenta* Crantz germplasm based on single-nucleotide polymorphism markers. **Annals of applied biology**, v. 173, n. 3, p. 271-284, 2018.

ALBUQUERQUE, H. Y. G. D.; OLIVEIRA, E. J. D.; BRITO, A. C.; ANDRADE, L. R. B. D.; CARMO, C. D. D.; MORGANTE, C. V.; FALEIRO, F. G. Identification of duplicates in cassava germplasm banks based on single-nucleotide polymorphisms (SNPs). **Scientia Agricola**, v. 76, p. 328-336, 2019.

ASARE, P. A.; GALYUON, I. K. A.; SARFO, J. K.; TETTEH, J. P. Morphological and molecular based diversity studies of some cassava (*Manihot esculenta* Crantz)

germplasm in Ghana. **African Journal of Biotechnology**, v. 10, n. 63, p. 13900-13908, 2011.

ASHA, K. I.; MARIA, V. A.; KOUNDINYA, A. V. V.; KRISHNAN, P. Molecular profiling of selected cassava (*Manihot esculenta* Crantz) germplasm using ISSR and SSR markers. **Journal of Root Crops**, v. 45, n. 2, p. 24-32, 2019.

BANGARWA, S. K.; BALWAN, K. L. S.; CHOUDHARY, M.; CHOUDHARY, M. K. An introduction to DNA-Markers and their role in crop improvement. **The Pharma Innovation Journal**. v. 7, p. 638-643. 2021.

BEUKELAER, H.; DAVENPORT, G. F.; FACK, V. Core Hunter 3: flexible core subset selection. **BMC bioinformatics**, v. 19, n. 1, p. 1-12, 2018.

BOARI, A. D. J., CUNHA, E. M., QUADROS, A. F. F., BARRETO, R. W., & FERNANDES, A. F. First report of *Phytophthium* sp. causing storage root rot and foliage blight of cassava in Brazil. **Plant disease**, v. 102, p. 1042-1042, 2018.

BORAH, B. K.; DASGUPTA, I. PCR-RFLP analysis indicates that recombination might be a common occurrence among the cassava infecting begomoviruses in India. **Virus genes**, v. 45, n. 2, p. 327-332, 2012.

BOCZKOWSKA, M.; ŁAPIŃSKI, B.; KORDULASIŃSKA, I.; DOSTATNY, D. F.; CZEMBOR, J. H. Promoting the use of common oat genetic resources through diversity analysis and core collection construction. **Plos one**, v. 11, n. 12, p. e0167855, 2016.

BOUKAR, O. BHATTACHARJEE, R.; FATOKUN, C.; KUMAR, P; GUEYE, B. Cowpea. In M. Singh, H.D. Upadhyaya and I.S. Bisht, Genetic and genomic resources of grain legumes improvement. **London: Elsevier**, p. 137-156, 2013.

BRADBURY, M. G.; EGAN, S.V.; BRADBURY, J.H. Determination of all forms of cyanogens in cassava roots and cassava products using picrate paper kits. **Journal Science Food Agricultural**, v. 79, p. 593-601, 1999.

BROWN, A. H. D. Core collections: a practical approach to genetic resources management. **Genome**, v. 31, n. 2, p. 818-824, 1989.

BROWN, A. H. D.; SPILLANE, C. Implementing core collections—principles, procedures, progress, problems and promise. In: Johnson RC, Hodgkin T (eds) Core Collections for Today and Tomorrow. **Crop Science Society of America**, Madison, p. 1-9, 1999.

BURLE, M. L.; OLIVEIRA, M. Manual de Curadores de Germoplasma-Vegetal: Caracterização Morfológica. **Brasília: Embrapa Recursos Genéticos e Biotecnologia**. p.15, 2010.

CARMO, C. D.; BRITO, A. C.; DE OLIVEIRA, E. J. Genome-wide association studies for waxy starch in cassava. **Euphytica**, v. 216, n. 5, p. 1-17, 2020.

CARMO, C. D. D.; SILVA, M. S. D.; OLIVEIRA, G. A. F.; OLIVEIRA, E. J. D. Molecular-assisted selection for resistance to cassava mosaic disease in *Manihot esculenta* Crantz. **Scientia Agricola**, v. 72, p. 520-527, 2015.

CARRASCO, N. F.; OLER, J. R. L.; MARCHETTI, F. F.; CARNIELLO, M.A.; AMOROZO, M. C. M.; VALLE, T.; VEASEY, E. A. Growing cassava (*Manihot esculenta*) in Mato Grosso, Brazil: genetic diversity conservation in small-scale agriculture. **Economic Botany**, v. 70, n. 1, p. 15-28, 2016.

CARVALHO, R. R. B.; MARMOLEJO CORTES, D. F.; BANDEIRA E SOUSA, M.; DE OLIVEIRA, L. A.; DE OLIVEIRA, E. J. Image-based phenotyping of cassava roots for diversity studies and carotenoids prediction. **PloS one**, v. 17, n. 1, p. e0263326, 2022.

CEBALLOS, H.; OKOGBENIN, E.; PÉREZ, J. C.; LÓPEZ-VALLE, L. A. B.; DEBOUCK, D. Cassava. In: **Root and tuber crops**. Springer, New York, NY. p.53-96, 2010

CHAIAREEKITWAT, S.; LATIF, S.; MAHAYOTHEE, B.; KHUWIJITJARU, P.; NAGLE, M.; AMAWAN, S.; MÜLLER, J. Protein composition, chlorophyll, carotenoids, and cyanide content of cassava leaves (*Manihot esculenta* Crantz) as influenced by cultivar, plant age, and leaf position. **Food Chemistry**, v. 372, p. 131-173, 2022.

COMPANHIA NACIONAL DE ABASTECIMENTO – **CONAB**, 2021 [acesso em outubro 2021]. Mandioca – Análise Mensal – setembro 2021 - Conab [online]. Disponível em: <https://www.conab.gov.br/>

CORDEIRO, C. M. T.; ABADIE, T. Coleções nucleares. **Recursos genéticos vegetais. Embrapa Recursos Genéticos e Biotecnologia**, Brasília, p. 575-604, 2007.

DANKWA, K. O.; PEPRAH, B. B. Industrialization of cassava sector in Ghana: progress and the role of developing high starch cassava varieties. **Ghana Journal of Agricultural Science**, v. 54, n. 2, p. 79-85, 2019.

JESUS BENEVIDES, C. M., SOUZA, M. V., SOUZA, R. D. B., & LOPES, M. V. Fatores antinutricionais em alimentos: revisão. **Segurança Alimentar e nutricional**, v. 18, n. 2, p. 67-79, 2011.

EL-SHARKAWY, M. A. Stress-tolerant cassava: the role of integrative ecophysiology-breeding research in crop improvement. **Open Journal of Soil Science**, v. 2, n. 02, p. 162, 2012.

ENGELMANN, F.; DUSSERT, S. Cryopreservation. In: Conservation of tropical plant species. Springer, New York, NY. p. 107-119, 2013.

FAOSTAT (Food and Agriculture Data). Produção, Área Colhida e produtividade de Mandioca no Mundo. Disponível em: <http://www.fao.org/faostat/en/#data/QC>. Acesso em 10 out. 2021.

FERGUSON, M. E.; SHAH, T.; KULAKOW, P.; CEBALLOS, H.A. Global overview



of cassava genetic diversity. **PLoS One**, v. 14, n. 11, p. e0224763, 2019.

FRANKEL, O.H., BROWN, A.H.D. Plant genetic resources today: a critical appraisal. In: HOLDEN, J.H.W., WILLIAMS, J.T. (Eds.), *Crop Genetic Resources: Conservation and Evaluation*. **Allen and Unwin, Winchester**, p. 249–257, 1984.

FUKUDA, W. M. G.; GUEVARA, C. L.; KAWUKI, R.; FERGUSON, M. E. Selected morphological and agronomic descriptors for the characterization of cassava. Ibadan, Nigeria: IITA, p.25, 2010.

FUKUDA, W. M. G.; COSTA, I. R. S.; SILVA, S. O. Manejo e conservação de recursos genéticos de mandioca (*Manihot esculenta* Crantz) na Embrapa Mandioca e Fruticultura Tropical. Cruz das Almas, BA: Embrapa - CNPMF, (**Embrapa – CNPMF. Circular técnica, 74**). p.4, 2005.

GARRIDO-CARDENAS, J. A.; MESA-VALLE, C.; MANZANO-AGUGLIARO, F. Trends in plant research using molecular markers. **Planta**, v. 247, n. 3, p. 543-557, 2018.

GOUESNARD, B.; BATAILLON, T. M.; DECOUX, G.; ROZALE, C.; SCHOEN, D. J.; DAVID, J. L. MSTRAT: An algorithm for building germ plasm core collections by maximizing allelic or phenotypic richness. **Journal of Heredity**, v. 92, n. 1, p. 93-94, 2001.

GROVER, A.; SHARMA, P. C. Development and use of molecular markers: past and present. **Critical reviews in biotechnology**, v. 36, n. 2, p. 290-302, 2014.

GUO, Y.; LI, Y.; HONG, H.; QIU, L. J. Establishment of the integrated applied core collection and its comparison with mini core collection in soybean (*Glycine max*). **The Crop Journal**, v. 2, n. 1, p. 38-45, 2014.

HOWELER, R.; LUTALADIO, N.; THOMAS, G. Save and grow: cassava. A guide to sustainable production intensification. **FAO**, p. 129, 2013.

IGE, A. D.; OLASANMI, B.; NKOUAYA MBANJO, E. G.; KAYONDO, I. S.; PARKES, E. Y.; KULAKOW, P.; RABBI, I. Y. Conversion and validation of uniplex SNP markers for selection of resistance to cassava mosaic disease in cassava breeding programs. **Agronomy**, v. 11, n. 3, p. 420, 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Levantamento Sistemático da Produção Agrícola. [Acesso em: 10 março 2022. Disponível em: <https://sidra.ibge.gov.br/tabela/1618>.

IRIANTO, H.; MUJIYO, M.; QONITA, A.; SULISTYO, A.; RIPTANTI, E. W. The development of jarak towo cassava as a high economical raw material in sustainability-based food processing industry. **AIMS Agriculture and Food**, v. 6, n. 1, p. 125-141, 2021.

KARIM, K. Y.; IFIE, B.; DZIDZIENYO, D.; DANQUAH, E. Y.; BLAY, E. T.; WHYTE, J. B.; NORMAN, P. E. Genetic characterization of cassava (*Manihot esculenta* Crantz) genotypes using agro-morphological and single nucleotide polymorphism markers. **Physiology and Molecular Biology of Plants**, v. 26, n. 2, p. 317-330, 2020.

KAWUKI, R. S.; HERSELMAN L.; LABUSCHAGNE M. T.; NZUKI I.; RALIMANANA I.; BIDIKA M. Genetic diversity of cassava (*Manihot esculenta* Crantz) landraces and cultivars from southern, eastern and central Africa. **Plant Genetic Resources**, v. 11, n. 2, p. 170-181, 2013.

KIM, K. W.; CHUNG, H. K.; CHO, G. T.; MA, K. H.; CHANDRABALAN, D.; GWAG, J. G.; PARK, Y. J. PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. **Bioinformatics**, v. 23, n. 16, p. 2155-2162, 2007.

KUHN, B. C.; MANGOLIN, C. A.; SOUTO, E. R.; VICIENT, C. M.; MACHADO, M. F. Development of retrotransposon-based markers IRAP and REMAP for cassava (*Manihot esculenta*). **Gene Conserve**, v. 15, n. 59, p. 01-10, 2016.

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

LESTARI, TRI; APRIYADI, Rion. Genetic potential of cassava biodiversity in Bangka Island, Indonesia. **Cell Biology and Development**, v. 1, n. 2, p. 41-45, 2017.

LIU, L.; LI, Y.; LI, S.; HU, N.; HE, Y.; PONG, R.; LIN, D. Comparison of next-generation sequencing systems. In: *The Role of Bioinformatics in Agriculture*. **Apple Academic Press**, p. 31-56. 2014.

LIU, J.; ZHENG, Q.; MA, Q.; GADIDASU, K. K.; ZHANG, P. Cassava Genetic Transformation and its Application in Breeding F. **Journal of integrative plant biology**, v. 53, n. 7, p. 552-569, 2011.

MA'ARUF, A. G.; ABDUL, H. R. Efficient processing of cassava starch: Physicochemical characterization at different processing parameters. **Food Research**, v. 4, p. 143-151, 2020.

MALIK, A. I.; KONGSIL, P.; NGUYỄN, V. A.; OU, W.; SREAN, P.; LÓPEZ-LAVALLE, L. A. B.; CEBALLOS, H. Cassava breeding and agronomy in Asia: 50 years of history and future directions. **Breeding Science**, p. 18180, 2020.

MAULANA, H.; KANAI, H. Multi-criteria decision analysis for determining potential agriculture commodities in Indonesia. **Journal of Engineering Science and Technology**, v. 15, p. 33-40, 2020.

MOURA, E. F.; FARIAS NETO, J. T.; SAMPAIO, J. E.; SILVA, D. T.; RAMALHO, G. F. Identification of duplicates of cassava accessions sampled on the North Region of Brazil using microsatellite markers. **Acta Amazonica**, v. 43, n. 4, p. 461-467. 2013.

NOIROT, M.; HAMON, S.; ANTHONY, F. The principal component scoring: a new method of constituting a core collection using quantitative data. **Genetic Resources and Crop Evolution**, v. 43, n. 1, p. 1-6, 1996.

NDJIONDJOP, M. N.; SEMAGN, K.; GOUDA, A. C.; KPEKI, S. B.; DRO TIA, D.; SOW, M.; WARBURTON, M. L. Genetic variation and population structure of *Oryza glaberrima* and development of a mini-core collection using DArTseq. **Frontiers in plant science**, v. 8, p. 1748, 2017

OCAMPO, J.; OVALLE, T.; LABARTA, R.; LE, D. P.; DE HAAN, S.; VU, N. A.; LOPEZ-LAVALLE, L. A. B. DNA fingerprinting reveals varietal composition of Vietnamese cassava germplasm (*Manihot esculenta* Crantz) from farmers' field and genebank collections. **Plant Molecular Biology**, p. 1-18, 2021.

ODONG, T. L.; JANSEN, J.; VAN EEUWIJK, F. A.; VAN HINTUM, T. J. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. **Theoretical and Applied Genetics**, v. 126, n. 2, p. 289-305, 2013.

OKOGBENIN, E.; SETTER, T. L.; FERGUSON, M.; MUTEKI, R.; CEBALLOS, H.; OLASANMI, B.; FREGENE, M. Phenotypic approaches to drought in cassava. **Frontiers in physiology**, v. 4, p. 93, 2013.

OLIVEIRA, E. J.; FERREIRA, C. F.; DA SILVA SANTOS, V.; DE JESUS, O. N.; OLIVEIRA, G. A. F.; DA SILVA, M. S. Potential of SNP markers for the characterization of Brazilian cassava germplasm. **Theoretical and Applied Genetics**, v. 127, n. 6, p. 1423-1440, 2014.

OLIVEIRA, E. J., FERREIRA, C. F., SANTOS, V. S., OLIVEIRA, G. A. Development of a cassava core collection based on single nucleotide polymorphism markers. **Genet Mol Res**, v. 13, n. 3, p. 6472-6485, 2014a.

OLIVEIRA, E.J.; OLIVEIRA FILHO, O. S.; SANTOS, V. S. Selection of the most informative morphoagronomic descriptors for cassava germplasm. **Pesquisa Agropecuária Brasileira**, v. 49, p. 891-900, 2014b.

OLIVEIRA, M. V.; BALIZA, D. P.; SOUZA, G. A.; CARVALHO, S. P.; ASSIS, L. H. Characterization of cassava clones using microsatellite markers. **Revista Ciência**

**Agronômica**, v. 43, n. 1, p. 170, 2012.

OLIVEIRA, P.H.G.A.; BARBOSA, A.C.O.; DINIZ, R.P. Morphological variation of starch granules in S1 cassava progenies. **Euphytica**, v. 214, p. 92, 2018.

OLIVEIRA, S. A. S.; BOAS, S. A. V.; BRAGANÇA, C. A. D.; OLIVEIRA, J. First report of *Phytophthora melonis* causing cassava wilt and root rot in the State of Bahia, Brazil. **Summa Phytopathologica**, v. 42, n. 1, p. 107, 2016b.

OLSEN, K. M. SNPs, SSRs and inferences on cassava's origin. **Plant Molecular Biology**, v. 56, n. 4, p. 517-526, 2004.

PERSEGUINI, J. M. K. C.; SILVA, G. M. B.; ROSA, J. R. B. F.; GAZAFFI, R.; MARÇAL, J. F.; CARBONELL, S. A. M.; BENCHIMOL-REIS, L. L. Developing a common bean core collection suitable for association mapping studies. **Genetics and molecular biology**, v. 38, p. 67-78, 2015.

PESSOA FILHO, M.; RANGEL, P. H. N.; FERREIRA, M. E. Extracting samples of high diversity from thematic collections of large gene banks using a genetic-distance based approach. **Genetics and molecular biology**, v. 10, n. 1, p. 1-10, 2010.

PLUMIER, G.; PANIS, B.; WILMS, H. Survival is insufficient, a cryopreservation case study on cassava. **Cuvilier Verlag**, p. 58-58. 2019.

SALOMÃO, A. N. Manual de Curadores de Germoplasma-Vegetal: Glossário. **Brasília, DF: Embrapa Recursos Genéticos e Biotecnologia**. p.14 2010.

SANTAYANA, M.; ROSSEL, G.; NÚÑEZ, J.; SØRENSEN, M.; DELÉTRE, M.; ROBLES, R.; HEIDER, B. Molecular characterization of cultivated species of the genus *Pachyrhizus* Rich. ex DC. by AFLP markers: Calling for more data. **Tropical Plant Biology**, v. 7, n. 3, p. 121-132, 2014.

SANTOS, C. S. D.; SOUSA, M. B.; BRITO, A. C.; DE OLIVEIRA, L. A.; CARVALHO, C. W. P.; DE OLIVEIRA, E. J. Genome-wide association study of cassava starch

paste properties. **PloS one**, v. 17, n. 1, p. e0262888, 2022.

SANTOS, P.P. E.; SOUSA, M.B.; OLIVEIRA, E. J. Prediction models and selection of agronomic and physiological traits for tolerance to water deficit in cassava. **Euphytica**, v. 215, p. 73, 2019.

SILVA, R. L.; FERREIRA, C. F.; DA SILVA LEDO, C. A.; DE SOUZA, E. H.; DA SILVA, P. H.; DE CARVALHO COSTA, M. A. P.; SOUZA, F. V. D. Viability and genetic stability of pineapple germplasm after 10 years of in vitro conservation. **Plant Cell, Tissue and Organ Culture**, v. 127, n. 1, p. 123-133, 2016.

SCHLÖTTERER, Christian. The evolution of molecular markers—just a matter of fashion. **Nature reviews genetics**, v. 5, n. 1, p. 63-69, 2004.

TAMBUN, Rut; MANIK, Yosef. Value chain analysis of cassava (*Manihot esculenta*) in Toba Samosir Regency, North Sumatera, Indonesia. **Journal of Engineering and Scientific Research**, v. 2, n. 1, p. 39-43, 2020.

TEIXEIRA, J. H.; GUIMARÃES, M. A. S.; CARDOSO, S. C.; DOS SANTOS BRITO, A.; DINIZ, R. P.; DE OLIVEIRA, E. J.; DE OLIVEIRA, S. A. S. Evaluation of resistance to bacterial blight in Brazilian cassava germplasm and disease-yield relationships. **Tropical Plant Pathology**, v. 46, n. 3, p. 324-335, 2021a.

TEIXEIRA, J. H. D. S.; GUIMARÃES, M. A. S.; CARDOSO, S. C.; BRITO, A. D. S.; PEIXOUTO, Y. S.; RIBEIRO, S. O.; ALVES SANTOS DE OLIVEIRA, S. Distribution of resistance of cassava genotypes to dry, soft and black root diseases and correlation to yield parameters. **Journal of Phytopathology**, v. 169, n. 6, p. 350-359, 2021.

UPADHYAYA, H. D.; GOWDA, C.L.L. **Managing and enhancing the use of germplasm—strategies and methodologies**. International Crops Research Institute for the Semi-Arid Tropics, n.10, p.236, 2009.

UPADHYAYA, H. D.; PUNDIR, R. P. S.; DWIVEDI, S. L.; GOWDA, C. L. L.; REDDY,

V. G.; SINGH, S. Developing a mini core collection of sorghum for diversified utilization of germplasm. **Crop Science**, v. 49, n. 5, p. 1769-1780, 2009a.

UPADHYAYA, H. D.; ORTIZ, R. A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. **Theoretical and Applied Genetics**, v. 102, n. 8, p. 1292-1298, 2001.

VAN HINTUM, T.J.L.; BROWN, A.H.D.; SPILLANE, C.; HODGKIN, T. Core Collection of Plant Genetic Resources. **IPGRI Technical Bulletin: International Plant Genetic Resources**, Roma, Italia, n. 3, ref. 75, p. 48, 2000.

VENTURINI, M. T.; SANTOS, L. R. dos; OLIVEIRA, E. J. Development of a diagrammatic scale for the evaluation of postharvest physiological deterioration in cassava roots. **Pesquisa Agropecuária Brasileira**, v.50, p.658-668, 2015.

VIDYA, P.; NAIR, A. G.; MOHAN, C.; VIJAYAN, A.; SREEKUMAR, J. Studying the Disease Resistance and Morphological Variations in Cassava. **International Journal of Bio-resource and Stress Management**, v. 10, n. 4, p. 436-444, 2019.

YOU, L.; WANG, Y.; ZHANG, T.; ZHU, Y.; REN, N.; JIANG, X.; ZHOU, Y. Genome-Wide Identification of Nitrate Transporter 2 (NRT2) Gene Family and Functional Analysis of MeNRT2. 2 in cassava (*Manihot esculenta* Crantz). **Gene**, v. 809, p. 146038, 2021.

ZAGO, B. W.; BARELLI, M. A. A.; HOOGERHEIDE, E. S. S.; CORRÊA, C. L.; DELFORNO, G. I. S.; DA SILVA, C. J. Morphological diversity of cassava accessions of the south-central mesoregion of the State of Mato Grosso, Brazil. **Embrapa Agrossilvipastoril**, v.3, n.16, p. 1-10,2017.

ZHOU, Q. N.; SUN, A. H.; LI, Z.; HUA, Y. W.; JIANG, Z. H.; HUANG, T. D.; HUANG, H. S. Cryopreservation and plant regeneration of anther callus in *Hevea* by vitrification. **African Journal of Biotechnology**, v. 11, n. 28, p. 7212-7217, 2012.

**CAPÍTULO 1****DESENVOLVIMENTO DE COLEÇÕES NUCLEARES DE  
MANDIOCA COM BASE EM CARACTERÍSTICAS MORFO-  
AGRONÔMICAS E MARCADORES MOLECULARES**



## Desenvolvimento de coleções nucleares de mandioca com base em características morfo-agronômicas e marcadores moleculares

### RESUMO

A mandioca (*Manihot esculenta* Crantz) é uma das espécies amiláceas mais importantes do mundo. O objetivo deste trabalho foi desenvolver coleções nucleares (CC) utilizando dados fenotípicos (15 descritores quantitativos e 33 qualitativos) além de dados genotípicos (20.023 marcadores *single-nucleotide polymorphism* – SNPs) em 1.486 acessos de mandioca. Seis CC foram obtidas utilizando duas estratégias de otimização baseadas em distâncias genéticas: AN - *Average accession-to-nearest-entry* e EN – *Average entry-to-nearest-entry* e suas combinações com dados fenotípicos e genotípicos. Parâmetros genéticos como diversidade genética (ISW), endogamia (*Fis*), heterozigosidade observada (*Ho*) e esperada (*Hs*) foram obtidas para avaliar a qualidade das CC. Não houve elevada coincidência na seleção dos acessos entre as seis coleções nucleares, por isso foi desenvolvida uma sétima coleção (CCons) contendo acessos selecionados por pelo menos duas CC. A maioria das coleções apresentaram parâmetros genéticos muito semelhantes à coleção completa, com exceção daquelas desenvolvidas pelo algoritmo AN. Por outro lado, as variações nos valores máximos e mínimos dos parâmetros *Ho*, *Hs* e *Fis* foram muito próximos à coleção completa. A CCons e aquela construída com dados genotípicos e algoritmo EN (GenEN) foram as que conseguiram reter o maior número de alelos (>97%). Embora não tenha havido diferença significativa (acima de 5%), a CCons foi a que apresentou o perfil de distribuição e média das características mais semelhantes à coleção completa, com algumas poucas exceções. Apenas o ISW das características qualitativas apresentou variações superiores à  $\pm 10\%$  quando comparado com a coleção completa. Com base na análise de componentes principais (PCA), a CCons selecionou acessos de mandioca com uma dispersão mais uniforme nos quatro quadrantes em comparação com as demais coleções. Nossos resultados demonstram o desenvolvimento de CC otimizadas e úteis para a melhoria eficiente dos programas de melhoramento e estudos de associação genômica.

**Palavras-chave:** *Manihot esculenta* Crantz, germoplasma, melhoramento, caracterização.

## Development of cassava core collections based on morphological and agronomic traits and SNPS markers

### ABSTRACT

Cassava (*Manihot esculenta* Crantz) is one of the most important starchy crop species in the world. The aim of this study was to develop core collections (CC) using both phenotypic data (15 quantitative and 33 qualitative descriptors) and genotypic data (20,023 single-nucleotide polymorphism – SNP) from 1,486 cassava accessions. We obtained six CC using two optimization strategies based on genetic distances: AN - average accession-to-nearest-entry and EN – average entry-to-nearest-entry, and its combinations with phenotypic and genotypic data. We also assessed genetic parameters such as genetic diversity (ISW), inbreeding (*F<sub>is</sub>*), observed (*H<sub>o</sub>*) and expected (*H<sub>s</sub>*) heterozygosity to evaluate the quality of the CC. The selection of accessions was not highly coincident among the six core collections, so we developed a seventh collection (CCons) containing accessions selected by at least two CC. Most collections had genetic parameters very similar to the complete collection, except for those developed by the AN algorithm. However, the variations in the maximum and minimum values of the *H<sub>o</sub>*, *H<sub>s</sub>* and *F<sub>is</sub>* parameters were very close to the complete collection. The CCons and the collection constructed with genotypic data and the EN algorithm (GenEN) retained the highest number of alleles (>97%). Although there was no significant difference (above 5%), the CCons had a distribution profile and trait's mean most similar to the complete collection, with a few exceptions. The ISW of qualitative traits was the only parameter that showed variations greater than  $\pm 10\%$  when compared to the complete collection. Based on principal component analysis (PCA), the CCons selected cassava accessions with a more uniform dispersion in the four quadrants compared to the other CC. These results demonstrate the development of optimized and useful CC for efficient breeding programs and genomic association studies.

**Keywords:** *Manihot esculenta* Crantz, germplasm, breeding, characterization.

## 1.0 INTRODUÇÃO

A mandioca (*Manihot esculenta* Crantz) possui um papel importante na segurança alimentar mundial pelo fato de ser consumida por milhares de pessoas em países da África, Ásia e América latina (LEBOT, 2009). É uma cultura economicamente importante com grandes demandas no mercado mundial e uma ampla variabilidade genética. Os primeiros registros formais de melhoramento de mandioca começaram em meados de 1930, no Instituto Agrônomo de Campinas em São Paulo, junto com a conservação *ex situ* de recursos genéticos vegetais (FUKUDA et al., 2002). Atualmente existem diversas coleções de germoplasma de mandioca pelo mundo que visam documentar, avaliar, preservar e disponibilizar a diversidade genética existente da espécie para os programas de melhoramento. No Brasil, a Empresa Brasileira de Pesquisa e Agropecuária — EMBRAPA mantém cerca de 4.000 acessos de mandioca conservados de forma *ex situ*, em campo e *in vitro* (HERSHEY, 2017).

O gerenciamento desses bancos é um grande desafio devido ao grande número de acessos e o alto custo de manutenção (VAN HINTUM et al., 2000). Problemas associados à sub-representação da diversidade presente em determinados biomas, a existência de duplicatas de acessos, uso limitado do germoplasma pelos usuários finais, regeneração insuficiente dos materiais conservados, a caracterização morfo-agronômica incompleta e o baixo investimento na coleta e manutenção dos recursos genéticos, também tem sido preocupante para o futuro da espécie (DÍEZ et al., 2018). Por outro lado, a demanda crescente pela conservação e preservação dos materiais genéticos, ocasiona o crescimento contínuo das coleções pela falta de metodologias rápidas de verificação da existência ou não de diversidade adicional dos novos germoplasmas.

A chave para o melhoramento genético de plantas bem-sucedidas é o acesso eficiente da variabilidade, e grandes coleções de germoplasma são apropriadas e desejáveis do ponto de vista da preservação da variabilidade genética (FRANKEL, BENNETT, 1970). Porém, existe um grande risco da sua utilidade e acessibilidade ser inversamente proporcional ao seu tamanho (FRANKEL et al., 1981). A principal consequência disso é o uso de apenas uma pequena fração da diversidade genética pelos programas de melhoramento.

Os bancos de germoplasma armazenam a variação genética necessária

para aumentos constantes de produtividade, resistências a estresses e qualidade nutricional no âmbito dos programas de melhoramento (WANG et al., 2017). Entretanto, as coleções de germoplasma de mandioca possuem milhares de genótipos (muitos deles duplicatas de acessos), que ainda não possuem caracterização e definição do seu potencial para uso como parental ou mesmo uso *per se* no sistema de produção comercial. Isso tem limitado o uso dos recursos genéticos de mandioca no aumento do potencial produtivo da cultura e sobretudo na mitigação de efeitos adversos oriundos das mudanças globais no clima.

Recentemente diferentes estratégias foram propostas para melhorar o manejo de grandes coleções, a exemplo das coleções nucleares, que se referem à seleção de conjuntos de amostras representativas da variabilidade genética de coleções com o mínimo de repetição (BROWN e SPILLANE, 1999). Devido ao seu menor tamanho, a coleção nuclear pode ser fenotipada de forma bastante completa para os principais descritores que definem os potenciais de uso no melhoramento da espécie. A diversidade morfo-agronômica e molecular da coleção completa representada por uma pequena porção de acessos detalhadamente caracterizados e avaliados pode incentivar pesquisadores e produtores a incorporarem novos germoplasmas nos programas de melhoramento e mesmo diretamente no sistema de produção (BOCZKOWSKA et al., 2016).

As coleções nucleares maximizam a diversidade e reduzem a duplicidade da coleção completa de forma a melhorar o manuseio e a eficiência da conservação e uso dos recursos genéticos da espécie. No caso da mandioca, o primeiro registro da formação de coleções nucleares foi feito pelo Centro Internacional de Agricultura Tropical (CIAT) (HERSHEY et al., 1994). Bhattacharjee et al. (2012) utilizaram 40 caracteres morfo-agronômicos avaliados em duas localidades com diferentes números de acessos em cada localidade, e 428 acessos foram selecionados com 90% da variação total para compor a coleção nuclear do International Institute of Tropical Agriculture - IITA. Oliveira et al. (2014) utilizaram 354 marcadores *single-nucleotide polymorphism* (SNP) para formação de coleções nucleares composta por diferentes números de acessos. No entanto, existem poucos relatos da formação de coleções nucleares de mandioca com maior cobertura genômica e integração de informações fenotípicas e genotípicas.

Embora a variação fenotípica seja essencial para as ações práticas de seleção nos programas de melhoramento genético, a criação de coleções

nucleares com base em um grande número de descritores morfológicos, agrônômicos avaliados em diferentes anos de cultivo e um grande número de marcadores, ainda não foi proposta para mandioca. Portanto, este trabalho teve como objetivo: i) desenvolver coleções nucleares de mandioca com base em dados descritores quantitativos, qualitativos e moleculares de forma isolada e conjunta, ii) avaliar a eficiência de diferentes métodos de seleção de coleções nucleares de mandioca na retenção de máxima diversidade genética, variância e demais parâmetros genéticos em comparação com a coleção completa, e iii) gerar uma coleção nuclear consolidada que represente a máxima variabilidade fenotípica e molecular dos genótipos de mandioca.

## **2.0 MATERIAL E MÉTODOS**

### **2.1 Material vegetal**

Foram avaliados 1.486 acessos pertencentes ao banco de germoplasma de mandioca (BAG) da Embrapa Mandioca e Fruticultura. Os acessos são provenientes de diferentes regiões brasileiras e de intercâmbio com países como Colômbia, Venezuela, Nigéria, entre outros. Os acessos de mandioca desta coleção são representados por variedades locais e melhoradas obtidas a partir de cruzamentos, seleção massal e/ou identificadas por produtores ou instituições de pesquisas.

Os ensaios de avaliação e caracterização do BAG foram implantados entre os anos de 2011 e 2021, nas cidades baianas de Cruz das Almas, Laje e Valença (Tabela S1). De acordo com a classificação de Köppen, o clima da região é do tipo Af (tropical, com temperatura média anual de 24,2°C, umidade de aproximadamente 80%, e com precipitação média anual de 1.300 mm), sendo que o período dos meses de março a julho é o mais chuvoso, e outubro e janeiro o mais seco. As informações detalhadas de tipo de solo, coordenadas geográficas e anos de avaliação de cada localidade estão apresentadas na Tabela S1.

### **2.2 Descritores morfo-agronômicos**

A caracterização dos acessos foi realizada utilizando descritores agro

morfológicos em escalas padronizadas, para folha, caule, raiz, flores e descritores agrônômicos divididos em qualitativos e quantitativos para a cultura da mandioca (Tabela 1), de acordo com Fukuda et al. (2010).

**Tabela 1.** Descritores qualitativos avaliados para a formação das coleções nucleares fenotípicas e genotípicas.

<b>Orgão da planta</b>	<b>Descritores</b>
Folha	<ul style="list-style-type: none"> <li>• Cor da folha apical (CFA)</li> <li>• Cor da folha desenvolvida (CFD)</li> <li>• Cor da nervura (CN)</li> <li>• Forma do lóbulo central (FLC)</li> <li>• Cor do pecíolo (CP)</li> <li>• Pubescência do broto apical (P)</li> <li>• Número de lóbulos (NL)</li> <li>• Posição do pecíolo (PP)</li> <li>• Sinuosidade das folhas (S)</li> <li>• Proeminência das cicatrizes das folhas (PCF)</li> </ul>
Caule	<ul style="list-style-type: none"> <li>• Comprimento da Estípula (CE)</li> <li>• Cor dos ramos terminais (CRT)</li> <li>• Comprimento da filotaxia (CF)</li> <li>• Ângulo de ramificação (AR)</li> <li>• Cor externa da casca do caule (CECC)</li> <li>• Cor do córtex do caule (CCC)</li> <li>• Cor da epiderme do caule (CEC)</li> <li>• Hábito de crescimento do caule (HCC)</li> <li>• Margem da estípula (ME)</li> <li>• Tipo de planta (TP)</li> <li>• Hábito de ramificação (HR)</li> <li>• Níveis de ramificação (NR)</li> </ul>
Raiz	<ul style="list-style-type: none"> <li>• Cor da polpa das raízes (CPR)</li> <li>• Cor do córtex das raízes (CCR)</li> <li>• Cor externa das raízes (CER)</li> <li>• Destaque do córtex das raízes (DCR)</li> <li>• Destaque da película das raízes (DPR)</li> <li>• Forma das raízes (FR)</li> <li>• Textura da epiderme das raízes (TER)</li> <li>• Posição das raízes (PR)</li> <li>• Presença de pedúnculo das raízes (PPR)</li> <li>• Constrições das raízes (CR)</li> </ul>
Flor	<ul style="list-style-type: none"> <li>• Presença de flores (PF)</li> </ul>

Os descritores quantitativos avaliados foram:

- Comprimento do lóbulo central (CLC), medido em centímetros a partir do ponto de inserção do lobo;
- Largura do lóbulo central (LLC), em centímetros, medido na parte mais larga do lobo;
- Relação entre o comprimento e largura do lóbulo central (RCL), em centímetros, 
$$RCL = \frac{\text{Comprimentodolobocentral}}{\text{Larguradolobocentral}};$$
- Comprimento do pecíolo (CP), medido em folhas do terço médio central da planta em centímetros;
- Espessura da entre casca das raízes (ECR), mensurando-se a epiderme junto com córtex em milímetro com auxílio de um paquímetro;
- Diâmetro médio das raízes (DMR), medido em milímetro na parte central das raízes;
- Comprimento média das raízes (CMR), medido em centímetros por todo comprimento das raízes;
- Conteúdo de ácido cianídrico nas raízes (HCN), determinado pelo método rápido do picrato, conforme Bradbury et al. (1999);
- Número de raízes por planta (NR);
- Altura de planta (AP), medido em centímetros;
- Teor de matéria seca (TMS), pelo método da secagem de 200 g de raízes em forno a 55°C até peso constante;
- Produtividade de matéria seca (PMS), mensurado pela multiplicação do teor de matéria seca pela produtividade de raízes ( $t/ha^{-1}$ );
- Produtividade total de raízes (PTR) ( $t/ha^{-1}$ );
- Produtividade da parte aérea (PPA) ( $t/ha^{-1}$ );
- Índice de colheita (IC) (Kg),  $Ic = \frac{\text{Pesototalderraízes}}{\text{Pesototaldaplanta}}$

### 2.3 Genotipagem dos acessos de mandioca

O DNA foi extraído de folhas jovens da mandioca utilizando o método CTAB (brometo de cetiltrimetilamônio), de acordo com Doyle e Doyle (1990), com algumas modificações como o aumento da concentração de 2-mercaptoetanol a 0,4% e adição de polivilpirrolidona (PVP). A qualidade do DNA foi avaliada em gel agarose

1% corado com brometo de etídio (1,0mg/L).

As amostras de DNA foram enviadas para genotipagem no *Genomic Diversity Facility da Cornell University* (<http://www.biotech.cornell.edu/brc/genomic-diversity-facility>), com o protocolo básico de *Genotyping-by-Sequencing* (GBS) (ELSHIRE et al., 2011). Inicialmente, as amostras foram digeridas pela enzima de restrição ApeKI, uma endonuclease de restrição do tipo II que reconhece uma sequência degenerada de 5 bases (GCWGC, onde W é A ou T) em comprimentos de 100 bp, (HAMBLIN, RABBI, 2014). Após a digestão das amostras e da ligação dos fragmentos de clivagem da ApeKI e do adaptador, foi realizado o sequenciamento em de um sistema multiplex com 192 amostras. A GBS foi realizada pela plataforma de genotipagem Genome Analyzer 2000 (Illumina, Inc., San Diego, CA). Os reads obtidos foram alinhados ao genoma de referência de mandioca v.6 (BREDESON et al., 2016) usando o software BWA (LI, DURBIN, 2009). Um total de 20.023 marcadores SNPs foram obtidos, estes marcadores estão distribuídos nos 18 cromossomos da mandioca. As análises das sequências e filtros de qualidade foram realizadas no software Tassel versão 5.2.37 (BRADBURY et al., 2007) com remoção dos alelos de frequência mínima (MAF) e dados perdidos (Call Rate) inferior a 5% e superior a 80%, respectivamente. Posteriormente, os demais dados perdidos foram imputados no software Beagle 4.1 (BROWNING, 2016).

## 2.4 Análise de dados

Como a maioria dos acessos foram caracterizados em diferentes ensaios entre 2011 e 2021, pequenas variações entre as classes dos descritores qualitativos foram observadas para alguns descritores. Neste caso, utilizou-se a moda como a classe de maior frequência das observações coletadas ao longo dos diversos anos, como medida de tendência de um determinado descritor para determinados acessos.

A análise do conjunto de dados quantitativos do banco ativo de germoplasma de mandioca foi realizada utilizando modelos lineares mistos, devido ao grau de desbalanceamento presente nos ensaios de avaliação do germoplasma de mandioca. O conjunto de dados contém informações sobre o ano e local de avaliação, os quais foram aqui denominados de ambientes. A análise conjunta



considerando todos os ambientes para cada descritor quantitativo foi realizada com base no seguinte modelo estatístico:  $y = Zg + Wb + Ti + e$ , no qual:  $y$  é o vetor de observações fenotípicas;  $g$  é o vetor de efeitos genotípicos, considerado como aleatório  $gN(0, \sigma_g^2)$ ;  $b$  é o vetor de efeitos alinhados de blocos dentro de ensaios, considerado como aleatório  $bN(0, \sigma_b^2)$ ;  $i$  é o vetor de efeitos da interação genótipos x ensaios, considerado como aleatório  $(0, \sigma_i^2)$ ;  $e$  é o vetor de efeitos de erros, considerado como aleatório  $eN(0, \sigma_e^2)$ .  $Z$ ,  $W$ ,  $T$  são as matrizes de incidência para os referidos efeitos, respectivamente. Esse modelo foi utilizado para obter as estimativas dos valores genéticos dos genótipos oriundos da avaliação de experimentos sob delineamento de blocos incompletos instalados em vários ensaios. As análises de modelo linear misto foram realizadas por meio do pacote *sommer* (COVARRUBIAS-PAZARAN, 2016) no ambiente R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022).

## 2.5 Definição das coleções nucleares

Foram obtidas seis coleções nucleares com base em critérios distintos baseados em distâncias entre os acessos da coleção completa e da coleção nuclear (Tabela 2). As coleções foram implementadas com o auxílio do pacote Core Hunter (BEUKELAER et al., 2018) no R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022) para seleção de 10% de acessos em relação ao tamanho da coleção completa. Para otimização das coleções foram utilizadas duas estratégias baseadas em distâncias genéticas: AN - *Average accession-to-nearest-entry* e EN - *Average entry-to-nearest-entry* descritas por Odong et al. (2013). No caso do critério AN, é calculada a média da distância entre os acessos da coleção completa e a entrada (acesso já incluído a coleção nuclear) mais próxima. Já no caso do EN, o objetivo é maximizar a distância média entre cada entrada e a entrada vizinha mais próxima na coleção completa, sendo que cada entrada deve ser o mais distinta possível uma da outra. Para cada método de otimização, diferentes conjuntos de dados foram utilizados (dados fenotípicos, dados genotípicos e dados fenotípicos + genotípicos de forma conjunta) com diferentes distâncias relacionadas.

Para as coleções obtidas somente com dados fenotípicos utilizou-se a matriz

de dissimilaridade de Gower (GOWER, 1971) por permitir a análise conjunta de variáveis numéricas e categóricas referentes aos dados quantitativos e qualitativos dos descritores avaliados. Para cada variável  $j$ , considera-se um coeficiente de semelhança,  $s_j$ , em um intervalo [0,1]. Comparando os elementos  $l$  e  $k$ , a similaridade entre os mesmos é dada por:  $d(l, k) = \left( \frac{\sum_{j=1}^{p+q} 1_j(l,k) s_j(l,k)}{\sum_{j=1}^{p+q} 1_j(l,k)} \right)$ . Em que  $1_j(l, k)$  é uma variável igual a 1 se  $l$  e  $k$  podem ser comparados a variável  $X_j$ .

Para as coleções obtidas somente com dados marcadores moleculares utilizamos a distância de Rodgers Modificada (WRIGHT, 1978),  $\left( MR_{ij} = \frac{1}{2L} \sqrt{\sum_{l=1}^L \sum_{a=1}^2 (p_{ila} - p_{jla})^2} \right)$ , sendo que  $L$  é o número total de marcadores,  $p_{ila}$  é a frequência alélica do alelo  $a$  do marcador  $l$  para o acesso  $i$ ,  $p_{jla}$  é a frequência alélica do alelo  $a$  do marcador  $l$  para o acesso  $j$ , e  $m_l$  é o número de alelos coincidentes para o marcador  $l$ .

**Tabela 2.** Resumo das estratégias utilizadas para obtenção das coleções nucleares de mandioca com base em 20.023 marcadores SNPs e 48 descritores agromorfológicos, divididos em qualitativos (33) e quantitativos (15).

Coleção	Dados de entrada	Otimização*	Distância
GenAN	SNPs	AN	Modified Rogers
GenEN		EN	
PhenAN	agromorfológicos	AN	Gower
PhenEN		AN	
GPmAN	SNPs + agromorfológicos	AN	Média Czekanowski (Manhattan) + Gower
GPmEN		EN	

GenAN – Coleção nuclear formada por dados genotípicos e estratégia de otimização AN - *Average accession-to-nearest-entry*; GenEN – Coleção nuclear formada por dados genotípicos e estratégia de otimização EN - *Average entry-to-nearest-entry*; PhenAN Coleção nuclear formada por dados fenotípicos e estratégia de otimização AN - *Average accession-to-nearest-entry*; PhenEN - Coleção nuclear formada por dados fenotípicos e estratégia de otimização AN - *Average accession-to-nearest-entry*; GPmAN – Coleção formada por dados morfoagronômicos + SNPs e estratégia de otimização AN - *Average accession-to-nearest-entry*; GPmEN - Coleção formada por dados morfoagronômicos + SNPs e estratégia de otimização EN - *Average entry-to-nearest-entry*.

Para a formação da coleção nuclear conjunta, os dados genotípicos e fenotípicos foram utilizados para obtenção de matrizes de distância (Gower e Manhattan para os dados fenotípicos e genotípicos, respectivamente). A distância

de Czekanowski (Manhattan) é calculada pela fórmula:  $d_{cz}(A, B) = \frac{1}{2L} \sum_{i=1}^L |x_i - y_i|$ , em que  $x_i$  e  $y_i$  são as frequências dos alelos no loci  $i$  para o indivíduo A e B, respectivamente;  $L$  é o número de loci  $x_i$  e  $y_i$ . A distância de Czekanowski foi implementada com o auxílio do pacote *dartR* (MIJANGOS et al. 2022) no R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022).

## 2.6 Análise de diversidade, comparação e validação dos métodos

A coincidência entre as coleções nucleares foi analisada com base no índice Kappa (COHEN, 1960) considerando os diferentes métodos de formação das coleções nucleares de mandioca. A coincidência dos acessos entre as coleções foi realizada usando um código binário, na qual os indivíduos selecionados e não selecionados receberam os códigos 1 e 0, respectivamente.

A diversidade genética nas coleções nucleares, coleção consolidada e coleção completa foi avaliada utilizando parâmetros de heterozigosidade observada ( $H_o = 1 - \sum k \sum i P_{kii}/np$ ); diversidade genética dentro da população -  $H_s = \tilde{n}/(\tilde{n} - 1) \left[ 1 - \sum_i P_i^{-2} - \frac{H_o}{2\tilde{n}} \right]$ ; e coeficiente de endogamia, ( $F_{is} = 1 - H_o/H_s$ ), implementados com o auxílio do pacote *hierfstat* (GOUDET, 2005) no R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022).

Conjuntamente a comparação entre as diferentes coleções nucleares, coleção consolidada a coleção completa para os dados fenotípicos foi realizada com base na dispersão das características quantitativas e qualitativas, e com base nos índices de diversidade de Shannon-Weaver para cada característica nas coleções completa e coleções individuais, de acordo com a seguinte fórmula  $H' = -\sum_{i=1}^n p_i \log_e p_i$ : em que  $p_i$  é a frequência observada da classe  $i$ , para a característica  $n$ ,  $n$  é o número das classes fenotípicas de uma característica,  $p_i$  é a frequência relativa e  $\log_e$  é o logaritmo natural. Todos os índices  $H'$  foram normalizados e divididos pela máxima,  $\log_e n$ , visando manter os valores entre 0 e 1 (monomorfismo até máxima diversidade fenotípica). Para as características qualitativas,  $k$  representou o número de classes ou notas do descritor, já para os caracteres quantitativos foram estimadas seis classes com os limites inferior e superior definidos pela amplitude total observada na coleção completa para cada característica (Tabela 3). Todas as análises foram realizadas no R versão 4.2.2 (R

DEVELOPMENT CORE TEAM, 2022).

**Tabela 3.** Estimação dos limites inferiores e superiores das classes fenotípicas a serem utilizadas para a estimação do índice de diversidade de Shannon-Weaver para as características quantitativas.

Classe fenotípica	Limite Inferior	Limite Superior
1	$min$	$\leq min + Amp/6$
2	$min + Amp/6$	$\leq min + 2xAmp/6$
3	$min + 2xAmp/6$	$\leq min + 3xAmp/6$
4	$min + 3xAmp/6$	$\leq min + 4xAmp/6$
5	$min + 4xAmp/6$	$\leq min + 5xAmp/6$
6	$min + 5xAmp/6$	$\leq max$

min: mínimo da característica da coleção completa, max: máximo da característica da coleção completa, Amp: amplitude da característica da coleção completa.'

A estrutura das coleções nucleares e consolidada em comparação com a coleção completa foi avaliada com base na análise de componentes principais (PCA) dos dados agromorfológicos via pacote *AMR* (BERENDS et al., 2022), e análise de componentes principais (PCA) dos dados moleculares foi implementada com auxílio do pacote *PCAtools* (BLIGHE, LUN 2020) ambos no R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022).

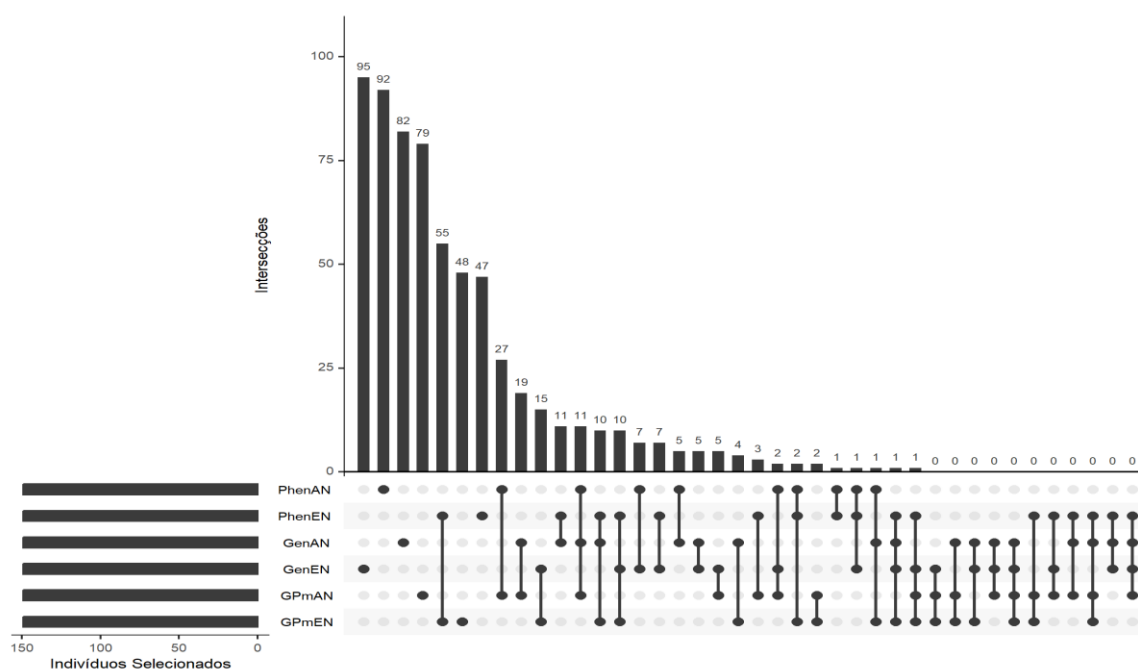
### 3.0 RESULTADOS

#### 3.1 Coincidência na seleção dos genótipos para formação das coleções nucleares com base em dados fenotípicos, genotípicos e dados agrupados

As coleções nucleares foram formadas considerando 10% da coleção completa, o que representa 149 genótipos. De modo geral, não houve elevada coincidência na seleção dos acessos entre as coleções nucleares. As coleções nucleares construídas com o algoritmo EN e baseada nos dados fenotípicos (PhenEN) e nos dados agrupados (fenotípicos e genotípicos) (GPmEN), foram as que tiveram o menor número de acessos exclusivos (ou seja, selecionados apenas nestas coleções): 47 e 48, respectivamente. Em contrapartida as demais coleções nucleares foram formadas com mais de 50% de acessos exclusivos e, portanto, baixa coincidência entre os acessos selecionados (Figura 1). Dos 1486 acessos da

coleção completa, 838 indivíduos não foram selecionados para compor nenhuma das coleções.

De acordo com o coeficiente Kappa, a coincidência de seleção dos acessos entre os diferentes métodos utilizados para formação das coleções nucleares variou de moderada (0.48) a baixa (-0.09) (Tabela 4). As maiores concordâncias na seleção dos genótipos de mandioca ocorreram nas coleções GPmEN e PhenEN (0.48) e entre GPmAN e PhenAN (0.19). Porém, a maior parte dos coeficientes foram negativos, o que demonstrou uma concordância pior do que a esperada ao acaso.



**Figura 1.** Número de acessos exclusivos e coincidentes considerando diferentes metodologias para formação de coleções nucleares: Coleções nucleares obtidas exclusivamente com dados fenotípicos e algoritmo de otimização *Average entry-to-nearest-entry distance* (PhenEN) e *Average accession-to-nearest-entry distance* (PhenAN); exclusivamente com dados genotípicos (GenAN e GenEN); e coleções agrupadas com ambos dados fenotípicos e genotípicos (GPmAN e GPmEN).

**Tabela 4.** Índice Kappa considerando diferentes metodologias para formação de coleções nucleares.

Índice Kappa	GenAN*	GenEN	PhenAN	PhenEN	GPmAN	GPmEN
<b>GenAN</b>	1	-0.07	0.02	0.05	0.11	0.01
<b>GenEN</b>	-0.07	1	-0.04	0.04	-0.05	0.09
<b>PhenAN</b>	0.02	-0.04	1	-0.08	0.19	-0.09
<b>PhenEN</b>	0.05	0.04	-0.08	1	-0.08	0.48
<b>GPmAN</b>	0.11	-0.05	0.19	-0.08	1	-0.09
<b>GPmEN</b>	0.01	0.09	-0.09	0.48	-0.09	1

\*GenAN – Coleção nuclear formada por dados genotípicos e estratégia de otimização AN - *Average accession-to-nearest-entry*; GenEN – Coleção nuclear formada por dados genotípicos e estratégia de otimização EN - *Averageentry-to-nearest-entry*; PhenAN Coleção nuclear formada por dados fenotípicos e estratégia de otimização AN - *Average accession-to-nearest-entry*; PhenEN - Coleção nuclear formada por dados fenotípicos e estratégia de otimização AN - *Average accession-to-nearest-entry*; GPmAN – Coleção formada por dados morfoagronômicos +SNPs e estratégia de otimização AN - *Average accession-to-nearest-entry*; GPmEN - Coleção formada por dados morfoagronômicos +SNPs e estratégia de otimização EN - *Averageentry-to-nearest-entry*.

### 3.2 Formação da coleção consolidada

Dos 1486 acessos de mandioca, 445 (~30%) foram selecionados por apenas um tipo de coleção nuclear, enquanto que a maioria dos acessos (~56%) não foram selecionados por nenhuma coleção (Tabela S2). Como as coleções nucleares possuem abordagens complementares para seleção dos acessos mais importantes para manutenção no germoplasma, e que há risco de não inclusão de acessos com alelos/características alternativas na coleção final, nós criamos uma sétima coleção nuclear, chamada de “coleção consolidada” que inclui acessos selecionados por pelo menos duas das abordagens utilizadas para obtenção das coleções anteriores. Com isso, a coleção consolidada foi composta por 204 acessos de mandioca (~14%) da coleção completa.

As coleções nucleares construídas com base no algoritmo EN nos dados fenotípicos (PhenEN) e dados agrupados (fenotípicos e genotípicos) (GPmEN), foram as que tiveram maior número de acessos compartilhados com a coleção consolidada (~50%). As demais coleções tiveram de 53 a 69 acessos que também foram selecionados pela coleção consolidada. As coleções que utilizaram dados agrupados construídas com o algoritmo EN e AN também tiveram um maior número de acessos coincidentes com a coleção consolidada (79 acessos) (Figura S1). As concordâncias do coeficiente Kappa entre a coleção consolidada e as coleções nucleares foram todas positivas havendo concordâncias de magnitude mediana

com as coleções PhenEN (0.44) e GPmEN (0.43) e baixa magnitude com as coleções GPmAN (0.20), GenAN (0.19), PhenAN (0.1) e GenEN (0.08) (Tabela S2).

### 3.3 Parâmetros genéticos com base nos dados moleculares da coleção completa, coleções nucleares e coleção consolidada

Com intuito de estimar a diversidade genética nas coleções nucleares, consolidada e coleção completa foram estimados parâmetros genéticos com base nos marcadores SNPs (Tabela 5). A maioria das coleções apresentaram parâmetros genéticos muito semelhantes em comparação com a coleção completa, exceto aquelas que utilizaram o algoritmo AN que foi capaz de manter a mesma  $H_o$  (0.403). Por outro lado, as variações nos valores máximos e mínimos para os três parâmetros ( $H_o$ ,  $H_s$  e  $F_{is}$ ) foram muito próximos a coleção completa. Em relação ao número total de alelo, verificou-se que as coleções CCons e GenEN foram as que conseguiram reter o maior número de alelos nas coleções nucleares (>97%).

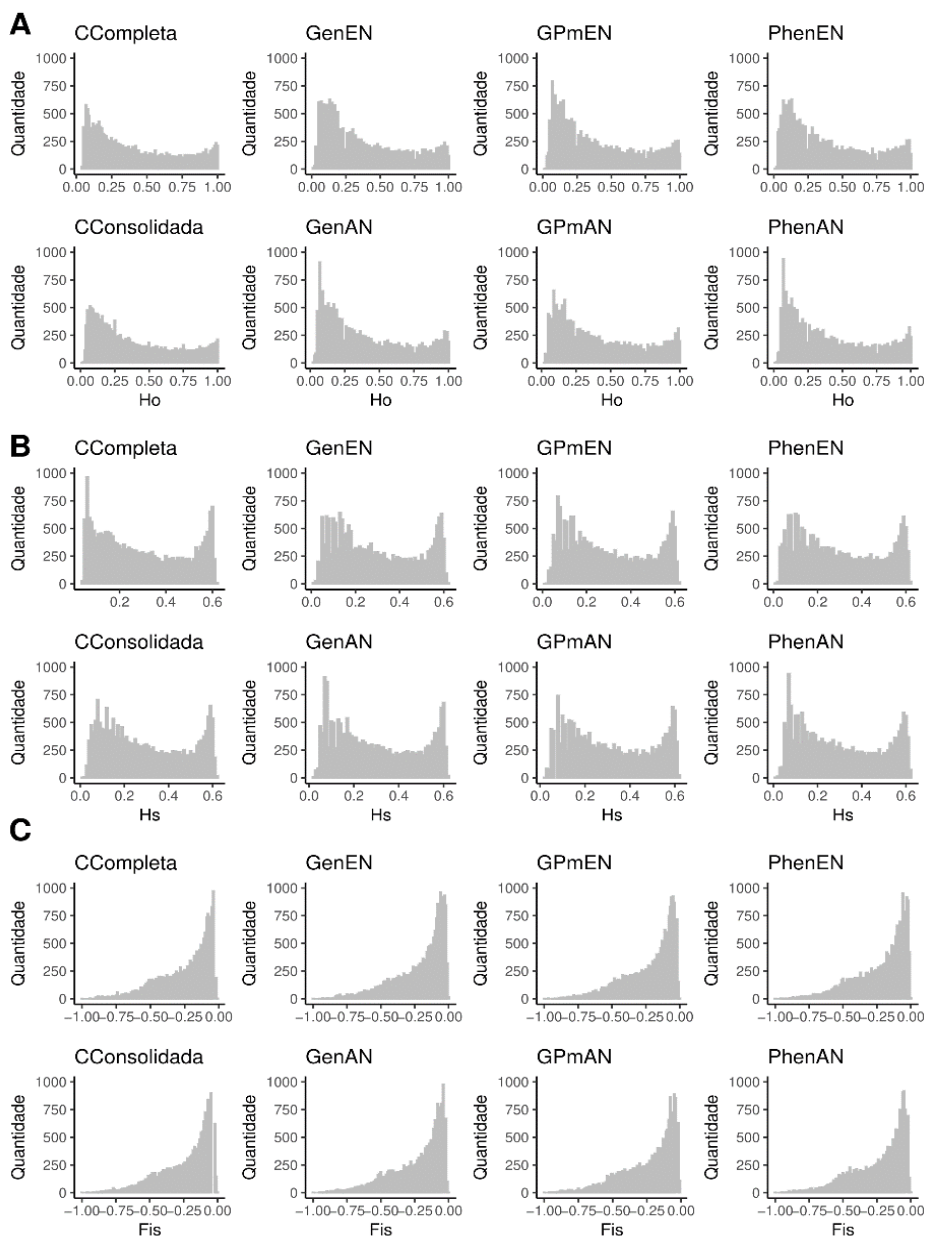
**Tabela 5.** Parâmetros básicos de diversidade genética para as coleções nucleares formadas com base em diferentes abordagens e informações de 20.023 marcadores SNPs.

Coleções	$H_o$		$H_s$		$F_{is}$		N# total de alelos
	Média	Variação	Média	Variação	Média	Variação	
Completa	0.403	(0.04 – 1.00)	0.301	(0.04 - 0.62)	-0.228	(-1.00/ -0.01)	58672
CCons	0.396	(0.01 – 1.00)	0.299	(0.01 – 0.62)	-0.220	(0.00/ -1.00)	56972
GenAN*	0.403	(0,02 – 1.00)	0.302	(0.02 – 0.62)	-0.226	(0.00/ -1.00)	55976
GenEN	0.390	(0.01 - 1.00)	0.296	(0.01 – 0.62)	-0.213	(0.00/ -1.00)	57338
PhenAN	0.403	(0.01 - 1.00)	0.302	(0.01 – 0.62)	-0.227	(0.00/ -1.00)	55464
PhenEN	0.396	(0.01 - 1.00)	0.298	(0.01 – 0.62)	-0.221	(0.00/ -1.00)	56021
GPmAN	0.403	(0.02 - 1.00)	0.302	(0.02 – 0.62)	-0.226	(-1.00/ -0.01)	55623
GPmEN	0.398	(0.01 - 1.00)	0.299	(0.01 – 0.62)	-0.220	(0.00/ -1.00)	56806

\*Coleções nucleares obtidas exclusivamente com dados fenotípicos e algoritmo de otimização Average entry-to-nearest-entry distance (PhenEN) e Average accession-to-nearest-entry distance (PhenAN); exclusivamente com dados genotípicos (GenAN e GenEN); coleções com ambos dados (GpmAN e GpmEN); Coleção consolidada que inclui acessos selecionados por pelo menos duas das abordagens utilizadas (CCons).  $H_o$  - Heterozigiosidade observada;  $H_s$  - diversidade genética dentro da população;  $F_{is}$  – Coeficiente de endogamia.

Em relação à distribuição dos parâmetros genéticos, verificou-se que os padrões das coleções nucleares foram relativamente parecidos com a coleção

completa (Figura 2). Algumas semelhantes mais evidentes foram identificadas entre a coleção consolidada e a completa para o parâmetro  $H_o$  e as coleções GenAn e PhenAN para o parâmetro  $H_s$ .

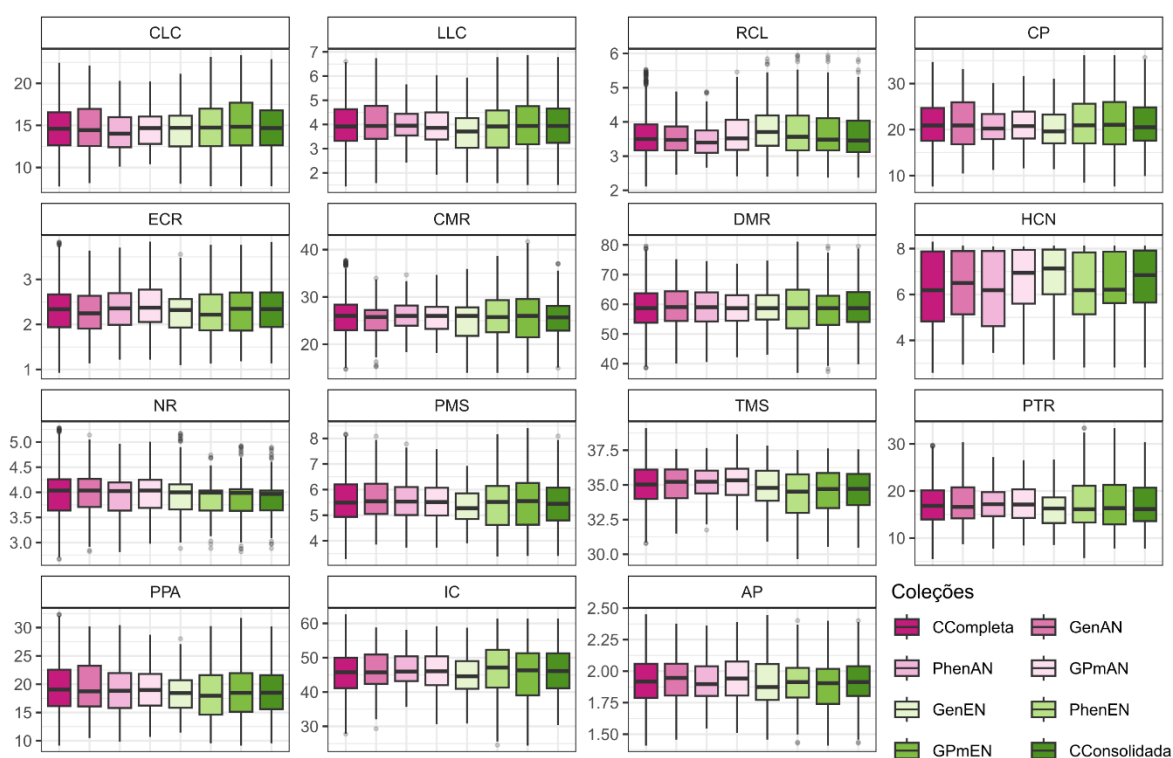


**Figura 2:** Parâmetros genéticos da coleção completa (CCompleta), coleção consolidada (CConsolidada) e coleções nucleares obtidas com base em dados fenotípicos (Phen) com uso do algoritmo *Average entry-to-nearest-entry distance* (AN) e *Average accession-to-nearest-entry distance* (EN) (PhenAN e PhenEN, respectivamente); dados genotípicos (GenAN e GenEN, respectivamente); dados fenotípicos e genotípicos agrupados (GPmAN e GPmEN, respectivamente). **A:** heterozigosidade observada ( $H_o$ ); **B:** heterozigosidade esperada ( $H_s$ ); **C:** coeficiente de endogamia ( $Fis$ ).



### 3.4 Variação nos descritores morfo-agronômicos da coleção completa, coleção consolidada e coleções nucleares

As faixas interquartis das características fenotípicas variaram em todas as coleções nucleares, embora as médias da maioria das características foram semelhantes à coleção completa em todas elas, exceto para as características relação comprimento de largura do lobo, conteúdo de ácido cianídrico nas raízes, espessura da entre casca das raízes, teor de matéria seca, altura de planta e índice de colheita, que apresentaram pequenas variações quando comparadas com a coleção completa (Figura 3). A coleção consolidada foi a que apresentou o perfil de distribuição e média das características mais semelhantes à coleção completa, com exceção do teor de ácido cianídrico nas raízes, número de raízes e índice de colheita.



**Figura 3.** Boxplot comparativo entre diferentes descritores quantitativos de mandioca (legenda em material e métodos) considerando diferentes coleções nucleares: Coleção completa (CCompleta), coleções nucleares obtidas com base em dados fenotípicos (Phen) com uso do algoritmo *Average entry-to-nearest-entry distance* (AN) e *Average accession-to-nearest-entry distance* (EM) (PhenAN e PhenEN, respectivamente); dados genotípicos (GenAN e GenEN, respectivamente); dados fenotípicos e genotípicos agrupadas (GPmAN e GPmEN, respectivamente) e coleção consolidada (CConsolidada).

As coleções nucleares obtidas com base nos dados fenotípicos, sejam isoladamente ou em conjunto com os dados genotípicos, apresentaram valores mínimos e máximos das características quantitativas muito próximos da coleção completa. Por outro lado, coleções baseadas em dados genotípicos, a exemplo da coleção GenAN apresentaram maior variação na média e amplitude dos dados fenotípicos, sobretudo para características relacionadas ao lóbulo da folha (comprimento do lóbulo central, largura do lóbulo central, relação comprimento e largura do lóbulo e comprimento do pecíolo). Acessos com valores extremos não foram selecionados para compor a coleção nuclear GenAN, enquanto que acessos com baixo índice de colheita também não foram incluídos nestas coleções nucleares.

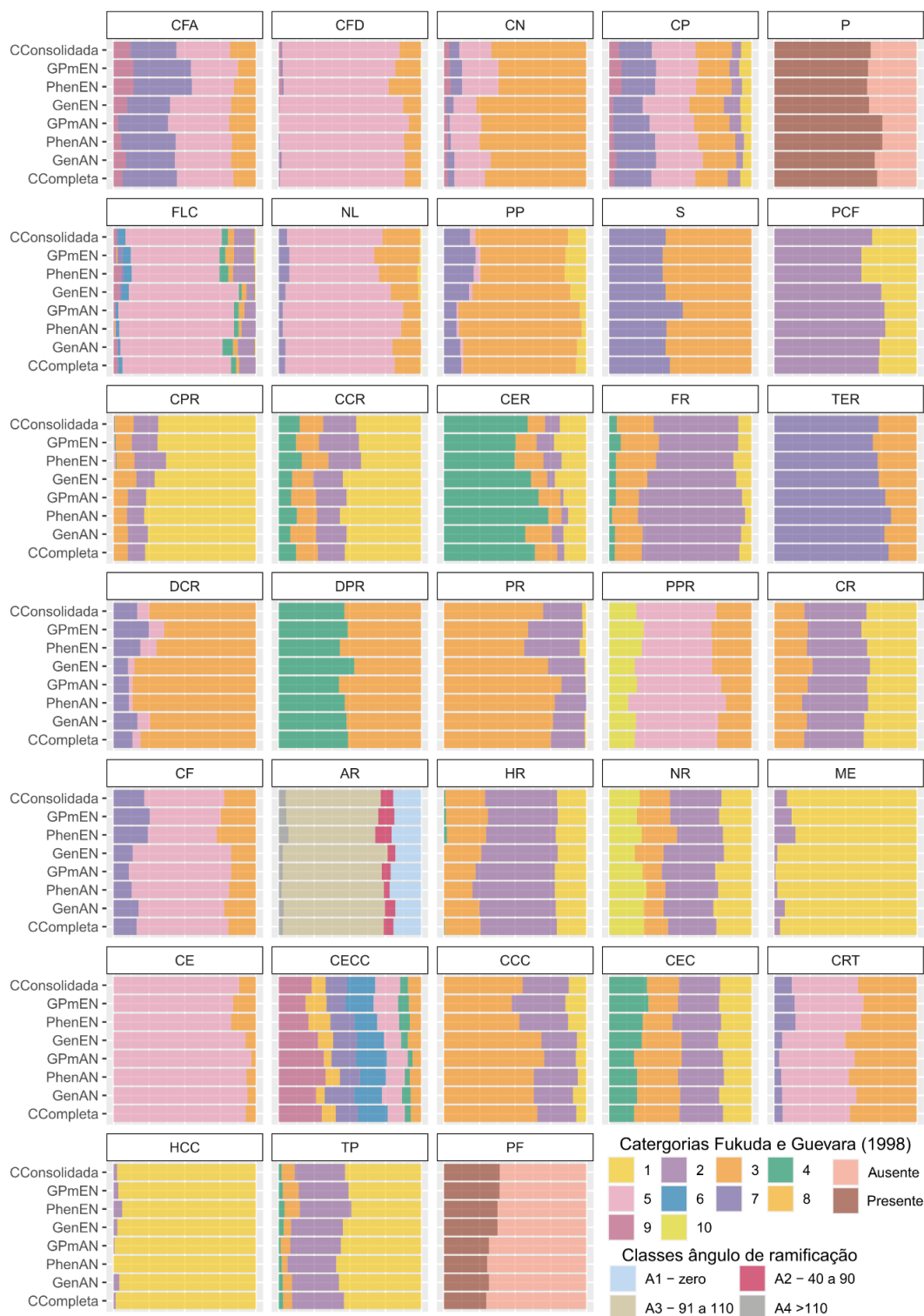
De modo geral, para a maioria das características fenotípicas quantitativas não houve diferença (acima de 5%) entre as médias e variâncias das coleções nucleares em comparação com a coleção completa (Tabela 6), embora as médias das características tenham apresentado alguma variação. Apenas a coleção GenAN apresentou médias mais elevadas em comparação com a coleção completa para a maioria das características, exceto relação comprimento e largura do lóbulo e espessura da entre casca das raízes. A variância foi superior à coleção completa (>50%) para algumas características como: i) diâmetro médio das raízes, em que as coleções GPmEN (125.90) e PhenEN (135.00), apresentaram variâncias consideravelmente superiores à coleção completa (79.31); e ii) índice de colheita, cujas coleções GPmEN (67.00) e PhenEN (66.19) também apresentaram variâncias superiores à coleção completa (45.95). Por outro lado, a coleção PhenAN apresentou variâncias inferiores à coleção completa em todas as características.

Os dados qualitativos foram analisados com base na eficiência das coleções nucleares em abranger todas as classes de cada uma das características avaliadas (Figura 4). As coleções nucleares que apresentaram um melhor equilíbrio na representação das classes foram CConsolidada, GPmEN e PhenEN, sobretudo para as características cor da nervura, número de lóbulos, posição do pecíolo, cor da polpa das raízes, posição das raízes, margem da estípula, habito de ramificação, ângulo de ramificação e cor externa da casca do caule.

**Tabela 6.** Média e variância para 15 características quantitativas avaliadas em diferentes coleções nucleares de mandioca formadas com base em dados fenotípicos (Phen), genótipos (Gen) e dados fenotípicos e genotípicos agrupadas (GPm) em 1.486 acessos de mandioca.

Característica	Col. completa		CCons		GenAN		GenEN		GPmAN		GPmEN		PhenAN		PhenEN	
	média	var	média	var	média	var	média	var	média	var	média	var	média	var	média	var
CLC	14,75	08,46	14.86	9.61	14,89	10,18	14,69	07,75	14,77	07,22	15,03	11,85	14,30	05,77	14,95	11,93
LLC	03,99	01,07	4.00	1.27	04,07	01,14	03,77	01,16	03,96	00,93	04,01	01,56	04,02	00,76	03,92	01,31
RCL	03,92	01,59	4.01	2.40	03,88	01,65	04,22	02,42	03,94	01,35	04,09	02,73	03,70	00,88	04,11	02,29
CP	21,19	26,16	21.31	32.20	21,34	34,19	20,28	21,88	21,34	20,56	21,40	40,06	20,90	20,47	21,17	40,48
EEC	02,39	01,04	2.36	0.41	02,33	00,41	02,38	00,52	02,45	00,36	02,47	02,91	02,37	00,35	02,46	04,21
CMR	26,17	30,17	25.92	38.31	26,21	35,64	25,60	30,10	25,81	21,29	26,06	42,85	26,24	17,39	25,96	42,51
DMR	58,96	79,31	59.57	103.95	59,33	72,60	58,35	61,18	59,82	70,80	58,69	125,90	59,02	66,55	59,30	135,0
HCN	06,27	02,53	6.49	2.40	06,38	02,50	06,66	02,17	06,55	02,53	06,44	02,36	06,20	02,46	06,33	02,48
NR	04,01	00,38	3.99	0.39	04,06	00,42	04,05	00,43	03,99	00,33	03,99	00,46	04,00	00,31	04,01	00,51
PMS	05,62	01,03	5.54	1.04	05,68	00,99	05,48	00,90	05,65	00,91	05,53	01,32	05,68	00,88	05,51	01,28
TMS	34,89	02,87	34.55	3.01	34,95	02,49	34,76	02,69	35,10	02,49	34,42	03,72	35,03	02,08	34,33	03,58
PTR	17,42	24,70	17.31	27.81	17,86	24,07	16,83	21,34	17,72	21,72	17,14	32,01	17,60	20,50	17,24	33,45
PPA	19,67	23,27	18.96	22.35	19,68	22,66	19,26	24,33	19,60	21,45	19,17	27,22	19,22	18,80	18,67	28,04
IC	45,28	45,95	45.82	56.41	45,54	49,48	44,76	46,58	45,76	45,09	45,34	67,00	46,31	31,43	46,21	66,19
AP	01,92	00,04	1.91	0.05	01,94	00,05	01,91	00,04	01,93	00,05	01,89	00,06	01,92	00,04	01,91	00,05

\*Coleções nucleares obtidas exclusivamente com dados fenotípicos e algoritmo de otimização *Average entry-to-nearest-entry distance* (PhenEN) e *Average accession-to-nearest-entry distance* (PhenAN); exclusivamente com dados genotípicos (GenAN e GenEN); coleções com ambos dados (GPmAN e GpmEN); Coleção consolidada que inclui acessos selecionados por pelo menos duas das abordagens utilizadas (CCons). \*CLC - Comprimento do lóbulo central; LLC - Largura do lóbulo central; RCL - Relação entre o comprimento e largura do lóbulo central; CP - Comprimento do pecíolo; EEC - Espessura da entre casca das raízes; DMR - Diâmetro médio das raízes; CMR - Comprimento médio das raízes; HCN - Conteúdo de ácido cianídrico nas raízes; NR - Número de raízes por planta; PP - Porte de planta; AP - Altura de planta; TMS - Teor de matéria seca; PMS - Produtividade de matéria seca; PR - Produtividade de raízes; PPA - Produtividade da parte aérea; IC - Índice de colheita; Var – Variância.



**Figura 4.** Barplot comparativo entre diferentes descritores qualitativos de mandioca (legenda em material e métodos) considerando diferentes coleções nucleares: Coleção completa, coleções nucleares obtidas com base em dados fenotípicos (Phen) com uso do algoritmo *Average entry-to-nearest-entry distance* (AN) e *Average accession-to-nearest-entry distance* (EM) (PhenAN e PhenEN, respectivamente); dados genotípicos (GenAN e GenEN, respectivamente), dados fenotípicos e genotípicos agrupadas (GPmAN e GPmEN, respectivamente) e coleção consolidada (CConsolidada).

Algumas características não foram bem representadas em todas as coleções nucleares possivelmente pelo fato de alguma das classes ser de ocorrência rara, sendo poucas vezes observada em campo. Por exemplo, apenas 28 acessos de mandioca (~1.88% da coleção completa) possuem hábito de crescimento do caule em zig-zag. Portanto, qualquer variação no método de formação da coleção nuclear pode alterar essa frequência, a exemplo da coleção PhenAN que foram representadas apenas por hábito de crescimento do tipo reto. Outras importantes oscilações na representatividade das coleções nucleares com base nos dados qualitativos foram identificadas para as características margem da estípula e posição das raízes também decorrente da baixa frequência de algumas classes na coleção completa.

### **3.5 Análise da diversidade fenotípica das coleções nucleares**

A qualidade das coleções nucleares foi avaliada com base no índice de Shannon-Weaver (ISW). Em comparação com a coleção completa, variações superiores à  $\pm 10\%$  do ISW para características fenotípicas quantitativas, foram observadas em poucas características (Tabela 7). Apenas o ISW da característica espessura da entre casca das raízes foi bastante afetado nas coleções nucleares, chegando a 0.00 nas coleções GenAN, PhenAN, GPmAN e CCons, e maior do que a coleção completa nas coleções GenEN, PhenEN e GPmEN. Em relação às demais características, a coleção GenAN foi a que apresentou menor diferença no ISW em relação à coleção completa, enquanto que a coleção PhenAN foi a que apresentou maior número de características com ISW inferior à coleção completa (comprimento do lóbulo central, relação entre comprimento e largura do lobo, comprimento do pecíolo, comprimento médio das raízes e índice de colheita). Por outro lado, as coleções PhenEN e GPmEN apresentaram maior número de características com ISW superior à coleção completa (relação entre comprimento e largura do lobo, comprimento do pecíolo, espessura da entre casca das raízes, comprimento e diâmetro das raízes e índice de colheita). A coleção consolidada foi a que apresentou maior número de características com menores diferenças em relação à coleção completa. Exceção ocorreu apenas para as características relação entre comprimento e largura do lobo, comprimento e diâmetro médio das raízes,

produtividade de matéria seca e índice de colheita, cujas diferenças do ISW foram acima de 5%.

**Tabela 7.** Índices de Shannon-Weaver obtidos com base em 15 descritores quantitativos avaliados em acessos de mandioca para desenvolvimento de seis coleções nucleares.

Caract.	Coleção							
	Completa	GenAN	GenEN	PhenAN	PhenEN	GPmAN	GPmEN	CCons
CLC	0,80	0,80	0,76	0,68	0,87	0,71	0,88	0,80
LLC	0,82	0,84	0,82	0,76	0,86	0,77	0,90	0,85
RCL	0,50	0,49	0,59	0,38	0,58	0,52	0,57	0,55
CP	0,70	0,75	0,65	0,62	0,79	0,64	0,80	0,72
EEC	0,01	0,00	0,02	0,00	0,02	0,00	0,02	0,00
CMR	0,46	0,49	0,50	0,34	0,52	0,41	0,52	0,51
DMR	0,59	0,53	0,53	0,56	0,72	0,55	0,68	0,64
HCN	0,86	0,86	0,79	0,85	0,86	0,83	0,83	0,84
NR	0,68	0,72	0,71	0,66	0,73	0,64	0,74	0,71
PMS	0,73	0,68	0,67	0,70	0,75	0,68	0,76	0,69
TMS	0,72	0,68	0,70	0,64	0,75	0,69	0,78	0,73
PTR	0,77	0,77	0,72	0,73	0,83	0,74	0,82	0,79
PPA	0,79	0,78	0,77	0,72	0,82	0,75	0,81	0,77
IC	0,77	0,77	0,78	0,67	0,86	0,74	0,87	0,83
AP	0,71	0,72	0,70	0,65	0,76	0,73	0,78	0,74

\*Coleções nucleares obtidas exclusivamente com dados fenotípicos e algoritmo de otimização *Average entry-to-nearest-entry distance* (PhenEN) e *Average accession-to-nearest-entry distance* (PhenAN); exclusivamente com dados genotípicos (GenAN e GenEN); e coleções com dados fenotípicos e genotípicos agrupados (GpmAN e GpmEN); Coleção consolidada que inclui acessos selecionados por pelo menos duas das abordagens utilizadas (CCons). \*CLC - Comprimento do lóbulo central; LLC - Largura do lóbulo central; RCL - Relação entre o comprimento e a largura do lóbulo central; CP - Comprimento do pecíolo; EEC - Espessura da entre casca das raízes; DMR - Diâmetro médio das raízes; CMR - Comprimento médio das raízes; HCN - Conteúdo de ácido cianídrico nas raízes; NR - Número de raízes por planta; ; AP - Altura de planta; TMS - Teor de matéria seca; PMS - Produtividade de matéria seca; PR - Produtividade de raízes; PPA - Produtividade da parte aérea; IC - Índice de colheita.

O ISW das características qualitativas apresentou variações superiores à  $\pm 10\%$  quando comparadas com a coleção completa, sobretudo nas coleções GenEN, PhenAN, PhenEN, GPmAN, GPmEN e CCons (Tabela 8). Similarmente às características quantitativas, a coleção GenAN foi a que apresentou menor variação do ISW para as características qualitativas, exceto habito de crescimento do caule, margem da estípula, cor externa das raízes e destaque do córtex das raízes que apresentaram maior ISW em comparação com a coleção completa e comprimento da estípula que apresentou redução do ISW. Nas

coleções nucleares PhenEN, GPmEN e CCons houve tendência para aumento do ISW em comparação com a coleção completa para a maioria das características (cor da folha desenvolvida, cor da nervura, forma do lóbulo central, número de lóbulos, posição do pecíolo, proeminência das cicatrizes das folhas, comprimento da estípula, cor dos ramos terminais, comprimento da filotaxia, ângulo de ramificação, cor do córtex do caule, habito de crescimento do caule, margem da estípula, cor da polpa das raízes, cor externa das raízes, destaque da película das raízes, forma das raízes, textura da epiderme das raízes e presença de flores), enquanto que nas coleções PhenAN e GPmAN houve tendência de redução do ISW, sobretudo para cor da nervura, forma do lóbulo central, número de lóbulos, posição do pecíolo, habito de crescimento do caule, margem da estípula, tipo de planta, cor externa das raízes, destaque do córtex das raízes, forma das raízes, posição das raízes e presença de pedúnculo das raízes.

**Tabela 8.** Índices de Shannon-Weaver obtidos com base em 33 descritores qualitativos de folha, caule raiz e flor, avaliados em acessos de mandioca para desenvolvimento de diferentes coleções nucleares.

Parte da planta	Caract.	Coleção							
		Completa	GenAN	GenEN	PhenAN	PhenEN	GPmAN	GPmEN	CCons
Folha	CFA	0,87	0,90	0,90	0,86	0,93	0,84	0,92	0,93
	CFD	0,26	0,29	0,26	0,25	0,43	0,18	0,40	0,34
	CN	0,58	0,61	0,51	0,51	0,72	0,51	0,71	0,66
	FLC	0,42	0,47	0,43	0,35	0,59	0,34	0,60	0,52
	CP	0,89	0,86	0,88	0,84	0,91	0,87	0,92	0,89
	P	0,85	0,88	0,92	0,80	0,93	0,80	0,93	0,91
	NL	0,43	0,42	0,48	0,34	0,58	0,33	0,57	0,52
	PP	0,50	0,50	0,66	0,39	0,77	0,40	0,75	0,71
	S	0,98	0,97	0,97	0,97	0,96	1,00	0,96	0,97
	PCF	0,84	0,83	0,81	0,76	0,96	0,78	0,96	0,90
Caule	CE	0,37	0,33	0,38	0,36	0,67	0,21	0,64	0,52
	CRT	0,79	0,78	0,79	0,77	0,92	0,74	0,91	0,96
	CF	0,82	0,86	0,76	0,76	0,96	0,71	0,95	0,90
	AR	0,61	0,62	0,57	0,56	0,76	0,61	0,72	0,69
	CECC	0,92	0,93	0,94	0,91	0,97	0,89	0,97	0,95
	CCC	0,59	0,63	0,57	0,59	0,70	0,56	0,72	0,69
	CEC	0,98	0,99	1,00	0,98	0,98	0,98	0,99	1,00
	HCC	0,12	0,24	0,18	0,00	0,33	0,06	0,21	0,17
	ME	0,24	0,38	0,14	0,14	0,60	0,06	0,53	0,43
	TP	0,71	0,69	0,70	0,64	0,78	0,67	0,76	0,72
	HR	0,75	0,73	0,73	0,70	0,80	0,72	0,79	0,77
	NR	0,98	0,97	0,95	0,96	0,99	0,96	0,98	0,98

**Tabela 8.** Cont...

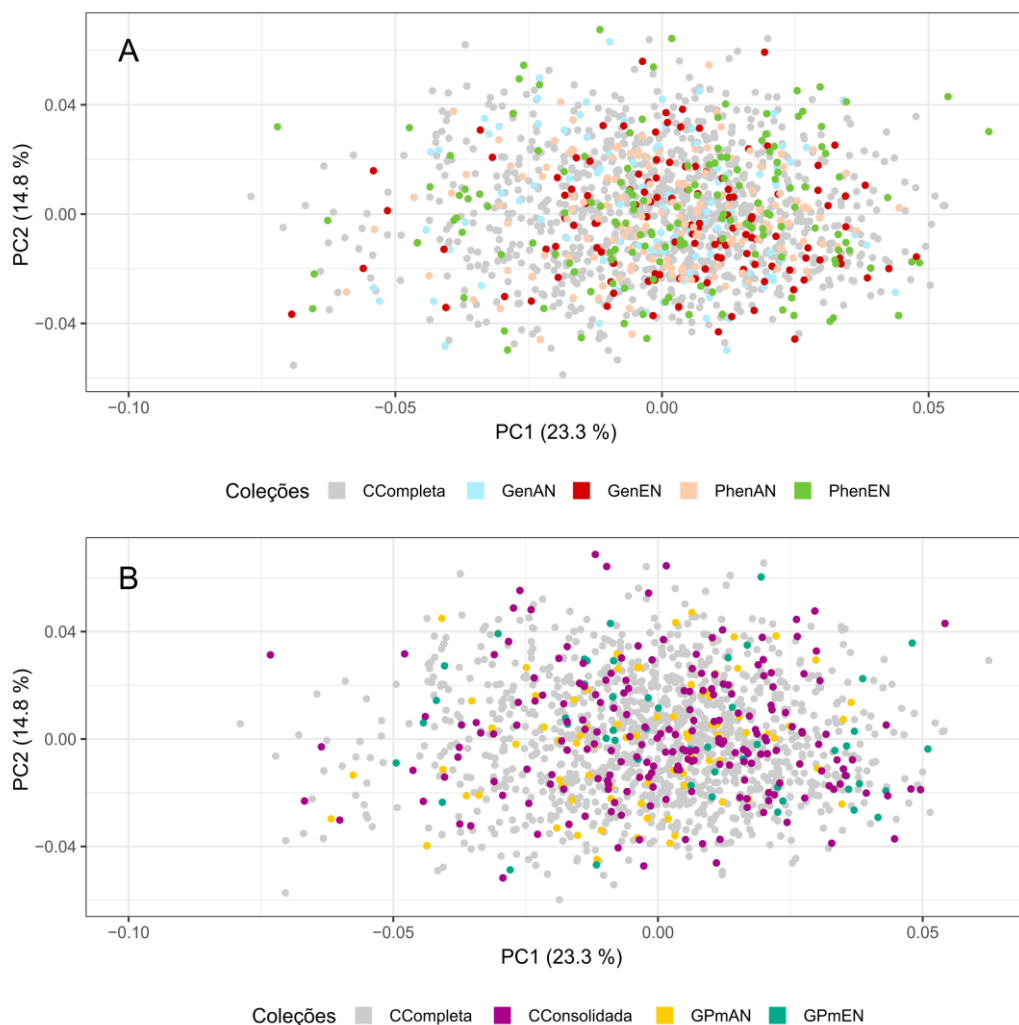
Parte da planta	Caract.	Coleção							
		Completa	GenAN	GenEN	PhenAN	PhenEN	GPmAN	GPmEN	CCons
Raiz	CPR	0,43	0,45	0,50	0,42	0,61	0,43	0,55	0,55
	CCR	0,86	0,85	0,84	0,83	0,95	0,86	0,92	0,92
	CER	0,73	0,82	0,75	0,62	0,89	0,67	0,88	0,81
	DCR	0,55	0,67	0,47	0,42	0,74	0,42	0,79	0,66
	DPR	1,00	1,00	1,00	1,00	0,99	0,98	1,00	1,00
	FR	0,66	0,65	0,73	0,54	0,79	0,62	0,80	0,75
	TER	0,72	0,78	0,84	0,68	0,85	0,76	0,89	0,84
	PR	0,55	0,54	0,58	0,48	0,76	0,45	0,71	0,65
	PPR	0,88	0,88	0,91	0,76	0,95	0,87	0,96	0,90
	CRT	0,97	0,98	0,99	0,95	0,97	0,98	0,98	0,89
Flor	PF	0,88	0,90	0,96	0,88	0,96	0,90	0,96	0,97

\*Coleções nucleares obtidas exclusivamente com dados fenotípicos e algoritmo de otimização *Average entry-to-nearest-entry distance* (PhenEN) e *Average accession-to-nearest-entry distance* (PhenAN); exclusivamente com dados genotípicos (GenAN e GenEN); e coleções com ambos dados (GPmAN e GPmEN); Coleção consolidada que inclui acessos selecionados por pelo menos duas das abordagens utilizadas (CCons). \*CFA – Cor da folha apical, CFD - Cor da folha desenvolvida, CN – Cor da nervura, FLC – Forma do lóbulo central, CP – Cor do pecíolo, P – Pubescência do broto apical, NL – Número de lobos, PP – posição do pecíolo, S – Sinuosidade das folhas, PCF – Proeminência das cicatrizes foliares, CE – Comprimento da estípula, CR – Constrictões das raízes, CF – Comprimento da filotaxia, AR – ângulo de ramificação. , CECC – Cor externa da casca do caule, CCC – Cor do córtex do caule, CEC – Cor da epiderme do caule, HCC – Hábito de crescimento do caule, ME – Margem da estípula, TP – Tipo de planta, HR – Hábito de ramificação, NR – Níveis de ramificação, CPR – Cor da polpa das raízes, CCR – Cor do córtex das raízes, CER – Cor externa das raízes, DCR – Destaque da película das raízes, DPR – Destaque da película das raízes, FR – Forma das raízes, TER – Textura da epiderme das raízes, PR – posição das raízes, PPR – Presença de pedúnculo das raízes, CRT – Cor dos ramos terminais, PF – Presença de flores.

### 3.6 Validação das coleções nucleares

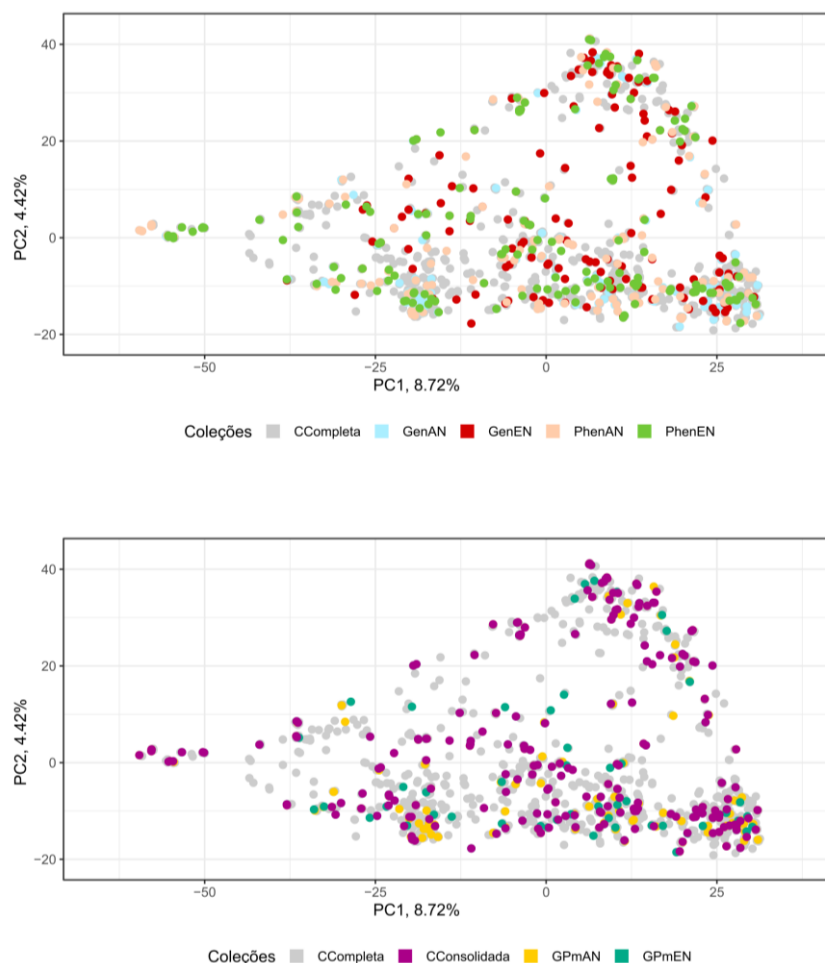
A PCA dos dados fenotípicos e moleculares foi utilizada para avaliar a representatividade da diversidade das coleções nucleares (Figuras 5 e 6, respetivamente). O primeiro e o segundo componente principal explicaram mais de 38% da variação fenotípica dos acessos de mandioca, tendo assim uma boa qualidade na representação da diversidade fenotípica dos acessos de mandioca (Figura 5). De modo geral, os acessos selecionados nas diferentes coleções nucleares ficaram bem distribuídos nos diferentes quadrantes da PCA dos dados fenotípicos. Entretanto, a coleção GPmEN selecionou maior número de acessos de mandioca distribuídos nos extremos dos quadrantes da PCA dos dados fenotípicos, enquanto a coleção consolidada selecionou acessos de mandioca com uma dispersão ligeiramente mais uniforme nos quatro quadrantes em comparação às demais coleções.





**Figura 5.** Análise de componentes principais (PCA) dos dados fenotípicos de 1486 acessos de mandioca com a dispersão das diferentes coleções nucleares. A: Coleção completa (CCompleta), coleções nucleares obtidas com base em dados fenotípicos (Phen) com uso do algoritmo Average entry-to-nearest-entry distance (AN) e Average accession-to-nearest-entry distance (EN) (PhenAN e PhenEN, respectivamente); dados genotípicos (GenAN e GenEN, respectivamente). B: dados fenotípicos e genotípicos agrupados (GPmAN e GPmEN, respectivamente) e Coleção consolidada que inclui acessos selecionados por pelo menos duas das abordagens utilizadas (Cconsolidada).

Na PCA dos SNPs os dois componentes principais explicaram uma menor porcentagem da variação molecular dos dados (8.72% e 4.42%, respectivamente) em comparação com os dados fenotípicos (Figura 6). Porém, de maneira análoga à PCA dos dados fenotípicos, todas as coleções nucleares foram bem representadas na PCA dos dados moleculares, embora a coleção consolidada e GPmEN tenham sido formadas com acessos presentes em todos os quadrantes e com uma distribuição mais representativa da coleção completa.



**Figura 6.** Análise de componentes principais (PCA) com base em 20.023 *single-nucleotide polymorphism* (SNP) de 1486 acessos de mandioca com a dispersão das diferentes coleções nucleares. A: Coleção completa (CCompleta), coleções nucleares obtidas com base em dados fenotípicos (Phen) com uso do algoritmo Average entry-to-nearest-entry distance (AN) e Average accession-to-nearest-entry distance (EN) (PhenAN e PhenEN, respectivamente); dados genotípicos (GenAN e GenEN, respectivamente). B: dados fenotípicos e genotípicos agrupados (GPmAN e GPmEN, respectivamente) e Coleção consolidada que inclui acessos selecionados por pelo menos duas das abordagens utilizadas (Cconsolidada).

## 4.0 DISCUSSÃO

### 4.1 Coincidência de seleção e diversidade dos dados fenotípicos e moleculares das coleções nucleares

Uma coleção nuclear é formada por um conjunto limitado de acessos derivados de coleções de germoplasma preexistentes que tem como objetivo

representar a máxima diversidade possível da coleção original (FRANKEL E BROWN, 1984). Usualmente recomenda-se o desenvolvimento de coleções nucleares que tenham pelo menos 10% do tamanho e no mínimo 70% da diversidade genética da coleção original (BROWN, 1989). Com base nessa recomendação, foram construídas diversas coleções nucleares de mandioca utilizando dados fenotípicos e genotípicos isoladamente e em combinação, além de uma coleção consolidada com base na retenção de acessos de mandioca selecionados por, pelo menos, duas coleções nucleares.

Embora as coleções nucleares tenham sido formadas de modo a maximizar a riqueza alélica e a variação fenotípica adequada para manutenção da variabilidade genética, a seleção dos acessos de mandioca com base nos dados fenotípicos e genotípicos não apresentou elevada coincidência. A baixa correlação entre dados morfológicos e moleculares, também foi observada em populações de batata (*Solanum tuberosum* L.) (BERDUGO-CELY et al., 2017). De acordo com estes autores, este fenômeno ocorre como resultado da seleção ao qual as populações estão submetidas, considerando que marcadores moleculares geralmente não são submetidos à seleção natural, enquanto os caracteres fenotípicos são submetidos à pressão de seleção e forte influência ambiental. Outra hipótese para explicar essa baixa coincidência de seleção é a baixa associação entre as regiões genômicas acessadas pelos SNPs com os caracteres fenotípicos avaliados (OLIVEIRA et al., 2012).

Diversas coleções nucleares foram construídas com uso de dados fenotípicos (UPADHYAYA et al., 2001; 2009; MAHMOODI et al., 2019). Embora, esse tipo de informação seja diretamente relacionado aos atributos agronômicos e produtivos, ainda existe a possibilidade de serem afetados por fatores ambientais e erros experimentais, além da interação genótipo x ambiente, o que faz com que as coleções nucleares construídas apenas com base no valor do fenótipo possam ser imprecisas em alguns casos (TANKSLEY; MCCOUCH, 1997). Por isso, diversos autores têm recomendado a formação de coleções nucleares que combinem estes dois tipos de dados para que seja possível obter uma coleção nuclear com a máxima representatividade da coleção original para os mais variados tipos de dados e características (KUMAR et al., 2016; XUE et al., 2021), sem que haja perdas de alelos úteis para ações de preservação e melhoramento da espécie.

Diante da baixa coincidência na seleção de acessos de mandioca entre as diferentes coleções nucleares e do risco de não inclusão de acessos que possuam características fenotípicas ou moleculares necessárias para manutenção e uso imediato ou futuro, foi construída a coleção nuclear consolidada baseada na escolha de acessos selecionados por, pelo menos, duas abordagens metodológicas (EN e AN) e diferentes tipos de coleções (Gen, Phen e GPm). Isso fez com o que o número de clones selecionados aumentasse um pouco (de 10% para ~14% da coleção completa), embora ainda perfeitamente manejável no âmbito dos programas de recursos genéticos e melhoramento da espécie.

De modo geral, todas as coleções nucleares de mandioca foram capazes de reter uma alta quantidade de alelos SNPs identificados na coleção completa (>94.5%), embora as coleções consolidadas e GenEN foram as que mantiveram a maior riqueza alélica (97.1% e 97.73%, respectivamente dos alelos da coleção completa). Essa manutenção da riqueza alélica nas coleções nucleares é de fundamental importância para estudos posteriores de associação genômica, sobretudo para características controladas por alelos raros na população. Além disso, nossos resultados foram bastante promissores considerando a alta riqueza alélica retida nas coleções nucleares de mandioca em comparação com outras espécies como milho (93% - TODOROVSKA et al., 2005) e tomate (92% - MARTINS et al., 2015).

Poucas alterações foram observadas nos parâmetros genéticos analisados, embora nossas coleções nucleares de mandioca construídas com base nas informações fenotípicas e genotípicas separadamente apresentaram maiores desvios nos valores de diversidade e parâmetros genéticos ( $H_o$ ,  $H_s$  e  $F_{is}$ ) em comparação com a coleção completa e consolidada. Em termos metodológicos, as coleções obtidas por meio dos algoritmos AN, apresentaram valores  $H_o$ ,  $H_s$  e  $F_{is}$  mais similares à coleção completa, possivelmente pelo fato deste algoritmo objetivar uma representação similar da distribuição das características da coleção completa na coleção nuclear. De modo geral, para marcadores bialélicos a variação de  $H_o$  (0,39 a 0,40) e  $H_s$  (0,29 a 0,30) identificada nas coleções nucleares é bastante semelhante a outros germoplasmas de mandioca a exemplo da coleção global de mandioca do IITA cujos valores de  $H_o$  variaram de 0,33 a 0,39 e  $H_s$  variaram de 0,31 a 0,34

(FERGUSON et al., 2019).

A distribuição e representatividade das coleções nucleares foram relativamente similares à coleção completa, levando-se em consideração as estimativas de média e variância dos dados fenotípicos quantitativos. Entretanto, o perfil de distribuição das médias e variâncias dos dados fenotípicos foi mais parecido com a coleção completa. Outros autores também reportaram poucas diferenças significativas entre a coleção completa e a coleção nuclear de variedade crioulas de milho (LI et al., 2005) e arroz (*Oryza sativa* L.) (YAN et al., 2009). Esses resultados, demonstram o elevado potencial dessas coleções nucleares em representar a coleção completa, sem resultar em perdas significativas de variabilidade genética.

Apenas em situações específicas, houve incremento na variância de algumas características, provavelmente em função da retirada de acessos com contribuição redundante para a diversidade fenotípica, cujos valores fenotípicos estão próximos à média amostral. Por exemplo, para a variância da característica diâmetro médio das raízes, as coleções consolidada, GPmEN e PhenEN, aumentou 31%, 49% e 62%, respectivamente, quando comparados com a coleção completa. Por outro lado, o aumento da variância nas coleções nucleares em comparação com a coleção original é um atributo desejável para haver uma maior representatividade da coleção completa com a menor redundância possível (HU et al., 2000), para seja possível obter coleções manejáveis para maior uso prático na descoberta de novas características e incorporação no melhoramento da espécie.

O ISW também foi utilizado para comparar a eficiência das coleções nucleares em representar a diversidade do germoplasma de mandioca. De modo geral, variações superiores à  $\pm 10\%$  do ISW em comparação com a coleção completa foram observadas para poucas características fenotípicas quantitativas. Apenas as características com maior desequilíbrio na sua distribuição foram mais afetadas quando ao ISW que chegou a 0.00 nas coleções obtidas com o algoritmo AN, a exemplo da espessura da entre casca das raízes. Por outro lado, as coleções obtidas com o algoritmo EN apresentaram ISW maior do que a coleção completa. Resultados semelhantes foram identificadas na análise dos dados fenotípicos qualitativos na qual as coleções com base no algoritmo EN e a coleção consolidada apresentaram maior ISW em comparação

com a coleção completa para a maioria das características. Entretanto, assim como em outras espécies de propagação clonal a exemplo do inhame (GIRMA et al., 2018; BEUKELAER et al., 2018), as coleções nucleares de mandioca mantiveram ou mesmo aumentaram a diversidade com base no ISW entre as coleções nucleares.

#### **4.2 Maximização da diversidade e representação das coleções nucleares**

Embora o intuito das coleções nucleares seja basicamente a maximização na representatividade da diversidade genética da coleção completa, seja com base em dados fenotípicos ou genotípicos, ainda existem diferenças metodológicas que podem influenciar na construção das coleções nucleares. As coleções nucleares deste estudo basearam-se em dois critérios descritos por Odong et al. (2013) que tem propósitos distintos. A coleção do tipo 1 (CC-I), visa representar toda a diversidade original de forma a maximizar a diversidade genética de toda coleção. Nesse tipo de coleção, características fenotípicas de baixa e alta frequência são representadas de maneira mais uniforme, a exemplo das coleções formadas pelo algoritmo AN que mantiveram a mesma heterozigosidade, diversidade genética dentro da população e coeficiente de endogamia mais próximos da coleção completa quando comparados com as demais coleções. O método AN, tem como objetivo minimizar a distância média dos acessos do conjunto de dados completo e o acesso selecionado mais próximo na coleção nuclear (que pode ser o próprio indivíduo).

A segunda coleção deste trabalho foi a do tipo 3 (CC-D), que objetiva representar a distribuição dos acessos da coleção completa, garantindo que a proporção selecionada da coleção completa reflita as contribuições numéricas das diferentes categorias na coleção nuclear. É o caso das coleções formadas pelo algoritmo EN que selecionaram acessos bem distribuídos, principalmente nos extremos dos diferentes quadrantes da PCA, proporcionando uma melhor representação de toda coleção. Isso favorece a obtenção de coleções mais diversificadas, onde cada indivíduo selecionado é suficientemente diferente do outro, formando assim subconjuntos com baixa redundância (ODONG et al., 2013) e elevada representatividade dos descritores utilizados para formação das

coleções. De fato, isso foi observado nas coleções que utilizaram esse objetivo de otimização, considerando a maior diversidade amostral com base no ISW.

A coleção consolidada foi construída para solucionar o problema da baixa coincidência na seleção dos acessos entre as coleções nucleares, como uma alternativa para melhor representar os acessos de mandioca entre as seis coleções desenvolvidas com base em diferentes tipos de dados. Portanto, a coleção consolidada foi eficiente não só em contornar a baixa coincidência, como também na melhoria da retenção de alelos, menor diferença nos parâmetros genéticos das coleções e na retenção da máxima diversidade no ISW em todas as características. Além de melhor representar as classes fenotípicas e genotípicas da coleção completa na PCA.

### **4.3 Validações das coleções nucleares**

A distribuição dos acessos selecionados nas diferentes coleções nucleares e consolidada de mandioca foi bastante representativa da coleção completa, de acordo com a análise PCA, tanto dos dados fenotípicos quanto dos dados moleculares. Mesmo com a presença de uma estrutura populacional, em ambos os tipos de dados, os acessos de mandioca foram bem distribuídos nos diferentes quadrantes da PCA, sobretudo nas coleções GPmEN e consolidada. Isso resultou na seleção de acessos de mandioca com menor nível de redundância nas coleções nucleares. Em outras espécies, a exemplo de *Lagenaria siceraria* (Mol.) Standl, a PCA das coleções nucleares construídas com base em diferentes tipos de dados fenotípicos foi capaz de representar a coleção completa com bastante precisão e ainda manter uma boa representação da distribuição geográfica dos acessos (WANG et al., 2021).

Não há proporção perfeita ou tamanho fixo para todas as coleções nucleares, pois espécies diferentes têm necessidades de pesquisa diferentes. Entretanto, a coleção consolidada descrita no presente estudo representa cerca de 14% da coleção completa e sua amplitude e complexidade dos recursos genéticos de mandioca parecem ter uma proporção amostral bastante adequada para fornecer uma referência útil e uma base para maior utilização dos recursos de germoplasma de mandioca e futuros programas de melhoramento.

#### **4.4 Aplicação das coleções nucleares na conservação, caracterização e uso dos recursos genéticos da mandioca.**

A conservação dos recursos genéticos de mandioca são fundamentais para fins de pesquisas e descoberta de genes de interesse agrônomo para uso nos programas de melhoramento genético de mandioca. Entretanto, Guo et al. (2014) reportaram o desafio de manter e utilizar a diversidade de acessos em um banco de germoplasma. Todas as etapas de conservação e caracterização são laboriosas, demandam tempo e recursos financeiros para sua execução. Nesse sentido, o desenvolvimento de coleções nucleares apresentadas neste trabalho é uma relevante alternativa para a redução dos custos dispendidos nas etapas de conservação e caracterização do germoplasma da cultura.

O objetivo principal para o desenvolvimento da coleção, além de reduzir o tamanho do conjunto e manter a representatividade genética, é definir prioridades na conservação, priorizar e definir esforços para caracterizações e avaliações, facilitar o acesso e incrementar o conhecimento da estrutura genética disponível nos bancos de germoplasma. A coleção nuclear consolidada, irá facilitar o manuseio e estudos de variabilidade e caracterização morfológicas e agrônomicas mais detalhadas. Todas estas medidas visam otimizar a conservação e uso do germoplasma da mandioca, garantindo a segurança dos recursos genéticos atualmente disponíveis. Além disso, essa coleção nuclear terá prioridade na conservação *in vitro* para que seja possível garantir a conservação livre de intemperes ambientais e ainda facilitar o rápido intercâmbio da coleção. Por outro lado, coleções genéticas devem ser dinâmicas e revisadas periodicamente para incorporar acessos adicionais, e com isso garantir que os genótipos mais valiosos sejam preservados e caracterizados para uso na conservação e no melhoramento da espécie.

#### **5.0 CONCLUSÃO**

Este estudo demonstrou a possibilidade de uso de diferentes abordagens metodológicas e tipos de dados para formação de coleções nucleares de mandioca de forma a manter a diversidade e parâmetros genéticos da coleção completa. Entretanto, a baixa coincidência na seleção dos acessos para



formação das coleções nucleares indicou a necessidade de construção de uma coleção alternativa (chamada de coleção consolidada) que foi construída incluindo acessos de mandioca selecionados por pelo menos dois algoritmos distintos de formação das coleções em combinação com os dados fenotípicos e genotípicos.

A coleção consolidada apresentou menor variação dos parâmetros genéticos analisados em comparação com a coleção completa, além de manter mais de 97% da riqueza alélica dos marcadores SNPs, e ainda combinar acessos selecionados com base em informações fenotípicas e genotípicas de forma conjunta. Ademais, a coleção consolidada apresentou dispersão dos dados e representação das classes nas características quantitativas e qualitativas bastante similar à coleção completa. Apesar de representar uma maior porcentagem da coleção completa do que o originalmente planejado (~14%), a coleção consolidada ainda possui um tamanho manejável para a otimização e maximização das ações de caracterização e uso deste germoplasma.

## 6.0 REFERÊNCIAS BIBLIOGRÁFICAS

BERDUGO-CELY, J.; VALBUENA, R. I.; SÁNCHEZ-BETANCOURT, E.; BARRERO, L. S.; YOCKTENG, R. Genetic diversity and association mapping in the Colombian Central Collection of *Solanum tuberosum* L. Andigenum group using SNPs markers. **PloS one**, v. 12, n. 3, p. e0173039, 2017.

BERENDS, M. S.; LUZ, C. F.; FRIEDRICH, A. W.; SINHA, B. N.; ALBERS, C. J.; GLASNER, C. AMR: an R package for working with antimicrobial resistance data. **Journal of Statistical Software**, v. 104, p. 1-31, 2022.

BEUKELAER, H.; DAVENPORT, G. Corehunter: multi-purpose core subset selection. R package version 3.2. 1. 2018.

BHATTACHARJEE, R.; DUMET, D.; ILONA, P.; FOLARIN, S.; FRANCO, J. Establishment of a cassava (*Manihot esculenta* Crantz) core collection based on agromorphological descriptors. **Plant Genetic Resources**, v. 10, p. 119-127,

2012.

BLIGHE, K.; LUN, A. PCAtools: everything principal components analysis. R package version, v. 2, n. 0, 2020.

BOCZKOWSKA, M.; ŁAPIŃSKI, B.; KORDULASIŃSKA, I.; DOSTATNY, D. F.; CZEMBOR, J. H. Promoting the use of common oat genetic resources through diversity analysis and core collection construction. **Plos one**, v. 11, n. 12, p. e0167855, 2016.

BRADBURY, M. G.; EGAN, S.V.; BRADBURY, J.H. Determination of all forms of cyanogens in cassava roots and cassava products using picrate paper kits. **Journal Science Food Agricultural**, v. 79, p. 593-601, 1999.

BRADBURY, P. J.; ZHANG, Z.; KROON, D. E.; CASSTEVENS, T. M.; RAMDOSS, Y.; BUCKLER, E. S. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, n. 19, p. 2633-2635, 2007.

BREDESON, J. V.; LYONS, J. B.; PROCHNIK, S. E.; WU, G. A.; HA, C. M.; EDSINGERGONZALES, E.; GRIMWOOD, J.; SCHMUTZ, J.; RABBI, I. Y.; EGESI, C.; NAULUVULA, P.; LEBOT, V.; NDUNGURU, J.; MKAMILO, G.; BART, R. S.; SETTER, T. L.; GLEADOW, R. M.; KULAKOW, P.; FERGUSON, M. E.; ROUNSLEY, S.; ROKHSAR, D. S. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. **Nature**, v. 34, p. 562-571, 2016.

BROWN, A. H. D. Core collections: a practical approach to genetic resources management. **Genome**, v. 31, n. 2, p. 818-824, 1989.

BROWN, A. H. D.; SPILLANE, C. Implement core collections – principles, procedures, progress, problems and promise. In: JOHNSON, R. C.; HODGKIN, T. **Core collections for today and tomorrow**. Roma: IPGRI, 1999.

BROWNING, B. L.; BROWNING, S. R. Genotype imputation with millions of reference samples. **The American Journal of Human Genetics**, v. 98, n. 1, p. 116-126, 2016.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and psychological measurement**, v. 20, n. 1, p. 37-46, 1960.

COVARRUBIAS-PAZARAN, G. Genome-assisted prediction of quantitative traits using the R package sommer. **PloS one**, v. 11, n. 6, p. e0156744, 2016.

DÍEZ, M. J.; DE LA ROSA, L.; MARTÍN, I.; GUASCH, L.; CARTEA, M. E.; MALLOR, C.; CASAÑAS, F. Plant genebanks: present situation and proposals for their improvement. The case of the Spanish network. **Frontiers in plant science**, v. 9, p. 1794, 2018.

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

ELSHIRE, R. J.; GLAUBITZ, J. C.; SUN, Q.; POLAND, J. A.; KAWAMOTO, K.; BUCKLER, E. S.; MITCHELL, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PloS one**, v. 6, n. 5, p. e19379, 2011.

FERGUSON, M. E.; SHAH, T.; KULAKOW, P.; CEBALLOS, H.A. Global overview of cassava genetic diversity. **PloS one**, v. 14, n. 11, p. e0224763, 2019.

FRANKEL, O. H.; BENNETT, E. Genetic resources in plants-their exploration and conservation. **Genetic resources in plants-their exploration and conservation**. 1970.

FRANKEL, O. H.; FRANKEL, O.; SOULÉ, M. E. Conservation and evolution. **CUP Archive**, 1981.

FRANKEL, O.H., BROWN, A.H.D. Plant genetic resources today: a critical appraisal. In: HOLDEN, J.H.W., WILLIAMS, J.T. (Eds.), *Crop Genetic Resources:*

Conservation and Evaluation. **Allen and Unwin, Winchester**, p. 249–257, 1984.

FUKUDA, W. M. G.; DE OLIVEIRA, S.; IGLESIAS, C. Cassava breeding. **Crop Breeding and Applied Biotechnology**, v. 2, n. 4, 2002.

FUKUDA, W. M. G.; GUEVARA, C. L.; KAWUKI, R.; FERGUSON, M. E. Selected morphological and agronomic descriptors for the characterization of cassava. **IITA**, 2010.

GIRMA, G.; BHATTACHARJEE, R.; LOPEZ-MONTES, A.; GUEYE, B.; OFODILE, S.; FRANCO, J.; ABBERTON, M. Re-defining the yam (*Dioscorea* spp.) core collection using morphological traits. **Plant Genetic Resources**, v. 16, n. 3, p. 193-200, 2018.

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

GOWER, J. C. A general coefficient of similarity and some of its properties. **Biometrics**, p. 857-871, 1971.

GUO, Y.; LI, Y.; HONG, H.; QIU, L. J. Establishment of the integrated applied core collection and its comparison with mini core collection in soybean (*Glycine max*). **The Crop Journal**, v. 2, n. 1, p. 38-45, 2014.

HAMBLIN, Martha T.; RABBI, Ismail Y. The effects of restriction-enzyme choice on properties of genotyping-by-sequencing libraries: A study in cassava (*Manihot esculenta*). **Crop Science**, v. 54, n. 6, p. 2603-2608, 2014.

HERSHEY, C. Research for development: The CIAT cassava program. 1994.

HERSHEY, C. Ex situ conservation of cassava genetic material. In: Achieving sustainable cultivation of cassava. **Burleigh Dodds Science Publishing**, v.2 p. 59-68. 2017.

HU, J.; ZHU, J.; XU, H. M. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. **Theoretical and Applied Genetics**, v. 101, p. 264-268, 2000.

KUMAR, S.; AMBREEN, H.; VARIATH, M. T.; RAO, A. R.; AGARWAL, M.; KUMAR, A.; JAGANNATH, A. Utilization of molecular, phenotypic, and geographical diversity to develop compact composite core collection in the oilseed crop, safflower (*Carthamus tinctorius* L.) through maximization strategy. **Frontiers in plant science**, v. 7, p. 1554, 2016.

LEBOT, V. Tropical root and tuber crops: cassava, sweet potato, yams and aroids. **Crop Production Science in Horticulture**, v. 2, p. 432, 2009.

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows–Wheeler transform. **Bioinformatics**, v. 25, n. 14, p. 1754-1760, 2009.

LI, Y.; SHI, Y.; CAO, Y.; WANG, T. Establishment of a core collection for maize germplasm preserved in Chinese National Genebank using geographic distribution and characterization data. **Genetic Resources and Crop Evolution**, v. 51, n. 8, p. 845-852, 2005.

MAHMOODI, R.; DADPOUR, M. R.; HASSANI, D.; ZEINALABEDINI, M.; VENDRAMIN, E.; MICALI, S.; NAHANDI, F. Z. Development of a core collection in Iranian walnut (*Juglans regia* L.) germplasm using the phenotypic diversity. **Scientia Horticulturae**, v. 249, p. 439-448, 2019.

MARTINS, F. A.; SILVA, D. J. H.; CARNEIRO, P. C. S. Establishment of a core collection based on the integration of morphoagronomic, phytopathological and molecular data. **Revista Ciência Agronômica**, v. 46, n. 4, p. 836-845, 2015.

MIJANGOS, J. L.; GRUBER, B.; BERRY, O.; PACIONI, C.; GEORGES, A. DartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. **Methods in Ecology and Evolution**, v. 13, n. 10, p. 2150-2158, 2022.

ODONG, T. L.; JANSEN, J.; VAN EEUWIJK, F. A.; VAN HINTUM, T. J. Quality of core collections for effective utilization of genetic resources review, discussion and interpretation. **Theoretical and Applied Genetics**, v. 126, n. 2, p. 289-305, 2013.

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

OLIVEIRA, E. J., FERREIRA, C. F., SANTOS, V. S., OLIVEIRA, G. A. Development of a cassava core collection based on single nucleotide polymorphism markers. **Genetics and Molecular Research**, v. 13, n. 3, p. 6472-6485, 2014.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing, reference index version 4.0.3. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available from, 2021.

TANKSLEY, S. D.; MCCOUCH, S. R. Seed banks and molecular maps: unlocking genetic potential from the wild. **science**, v. 277, n. 5329, p. 1063-1066, 1997.

TODOROVSKA, E.; ABUMHADI, N.; KAMENAROVA, K.; ZHELEVA, D.; KOSTOVA, A.; CHRISTOV, N.; ATANASSOV, A. Biotechnological Approaches for Cereal Crops Improvement: Part II: Use of Molecular Markers in Cereal Breeding. **Biotechnology & Biotechnological Equipment**, v. 19, n. sup3, p. 91-104, 2005.

UPADHYAYA, H. D.; ORTIZ, R. A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. **Theoretical and Applied Genetics**, v. 102, p. 1292-1298, 2001.

UPADHYAYA, H. D.; PUNDIR, R. P. S.; DWIVEDI, S. L.; GOWDA, C. L. L.; REDDY, V. G.; SINGH, S. Developing a mini core collection of sorghum for

diversified utilization of germplasm. **Crop Science**, v. 49, n. 5, p. 1769-1780, 2009.

VAN HINTUM, T.J.L.; BROWN, A.H.D.; SPILLANE, C.; HODGKIN, T. Core collection of plant genetic resources. **IPGRI Technical Bulletin: International Plant Genetic Resources**, Roma, Italia, n. 3, ref. 75, p. 48, 2000.

WANG, Y.; WU, X.; LI, Y.; FENG, Z.; MU, Z.; WANG, J.; LI, G. Identification and validation of a core single-nucleotide polymorphism marker set for genetic diversity assessment, fingerprinting identification, and core collection development in bottle gourd. **Frontiers in Plant Science**, v. 12, p. 747940, 2021.

WANG, C.; HU, S.; GARDNER, C.; LÜBBERSTEDT, T. Emerging avenues for utilization of exotic germplasm. **Trends in Plant Science**, v. 22, n. 7, p. 624-637, 2017.

WRIGHT, S. Evolution and the genetics of populations: A treatise in four volumes Volume IV. University of Chicago Press; 1978.

XUE, H.; YU, X.; FU, P.; LIU, B.; ZHANG, S.; LI, J.; MA, W. Construction of the Core Collection of *Catalpa fargesii* f. *duclouxii* (Huangxinzimu) Based on Molecular Markers and Phenotypic Traits. **Forests**, v. 12, n. 11, p. 1518, 2021.

YAN, W. G.; LI, Y.; AGRAMA, H. A.; LUO, D., GAO, F.; LU, X.; REN, G. Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). **Molecular breeding**, v. 24, n. 3, p. 277-292, 2009.

## ANEXOS

**Tabela S1.** Localização geográfica e caracterização de campos dos ensaios do Banco Ativo de germoplasma de Mandioca.

Ano	Delineamento	Localização	Cidade	Altitude	Coordenadas	Tipo de solo
2011	DBC	Embrapa	Cruz das Almas	215	12°40'36.7"S, 39°05'08.0"W	LA
	DBC	Embrapa	Cruz das Almas	210	12°40'32.6"S, 39°05'13.7"W	LA
2012	DBA	Embrapa	Cruz das Almas	215	12°40'36.7"S, 39°05'08.0"W	LA
	DBA	Embrapa	Cruz das Almas	210	12°40'32.6"S, 39°05'13.7"W	LA
	DBA	Embrapa	Cruz das Almas	200	12°40'47.4"S, 39°05'00.2"W	LA
2013	DBC	Coopamido	Laje	180	13°06'38.4"S, 39°16'20.4"W	LV
	DBC	UFRB	Cruz das Almas	210	12°39'25.9"S, 39°04'58.8"W	LA
	DBC	Embrapa	Cruz das Almas	210	12°40'32.6"S, 39°05'13.7"W	LA
	DBC	Embrapa	Cruz das Almas	217	12°40'22.8"S, 39°05'06.1"W	LA
	DBA	Embrapa	Cruz das Almas	216	12°40'22.8"S, 39°05'01.5"W	LA
2014	DBA	Embrapa	Cruz das Almas	216	12°40'22.8"S, 39°05'01.5"W	LA
	DBA	Coopamido	Laje	180	13°06'39.6"S, 39°16'17.6"W	LA
	DBA	Coopamido	Laje	175	13°06'35.6"S, 39°16'19.3"W	LA
	DBA	Embrapa	Cruz das Almas	216	12°40'22.8"S, 39°05'01.5"W	LA
	DBC	Coopamido	Laje	180	13°06'39.6"S, 39°16'17.6"W	LA
2015	DBA	Embrapa	Cruz das Almas	216	12°40'22.8"S, 39°05'01.5"W	LA
	DBA	Bahiamido	Laje	196	13°06'38.5"S, 39°16'49.0"W	LA
	DBA	Bahiamido	Valença	40	13°15'33.5"S, 39°14'12.8"W	LVA
	DBA	UFRB	Cruz das Almas	210	12°39'16.4"S, 39°04'53.4"W	LA
	DBA	Bahiamido	Laje	296	13°06'38.5"S, 39°16'49.0"W	LA
2016	DBA	Embrapa	Cruz das Almas	216	12°40'19.5"S, 39°05'02.5"W	LA
	DBA	Embrapa	Cruz das Almas	216	12°40'19.5"S, 39°05'02.5"W	LA
	DBA	Embrapa	Cruz das Almas	216	12°40'19.5"S, 39°05'02.5"W	LA
2017	DBA	Embrapa	Cruz das Almas	216	12°40'19.5"S, 39°05'02.5"W	LA
2018	DBC	UFRB	Cruz das Almas	223	12°39'51.4"S, 39°04'15.7"W	LA
	DBC	UFRB	Cruz das Almas	223	12°39'43.5"S, 39°04'12.0"W	LA
2019	DBC	UFRB	Cruz das Almas	223	12°39'43.5"S, 39°04'12.0"W	LA
	DBC	UFRB	Cruz das Almas	223	12°39'43.5"S, 39°04'12.0"W	LA
2020	DBC	UFRB	Cruz das Almas	223	12°39'49.2"S, 39°03'58.1"W	LA
	DBC	UFRB	Cruz das Almas	225	12°39'49.2"S, 39°03'58.1"W	LA
2021	DBA	UFRB	Cruz das Almas	225	12°39'49.2"S, 39°03'58.1"W	LA

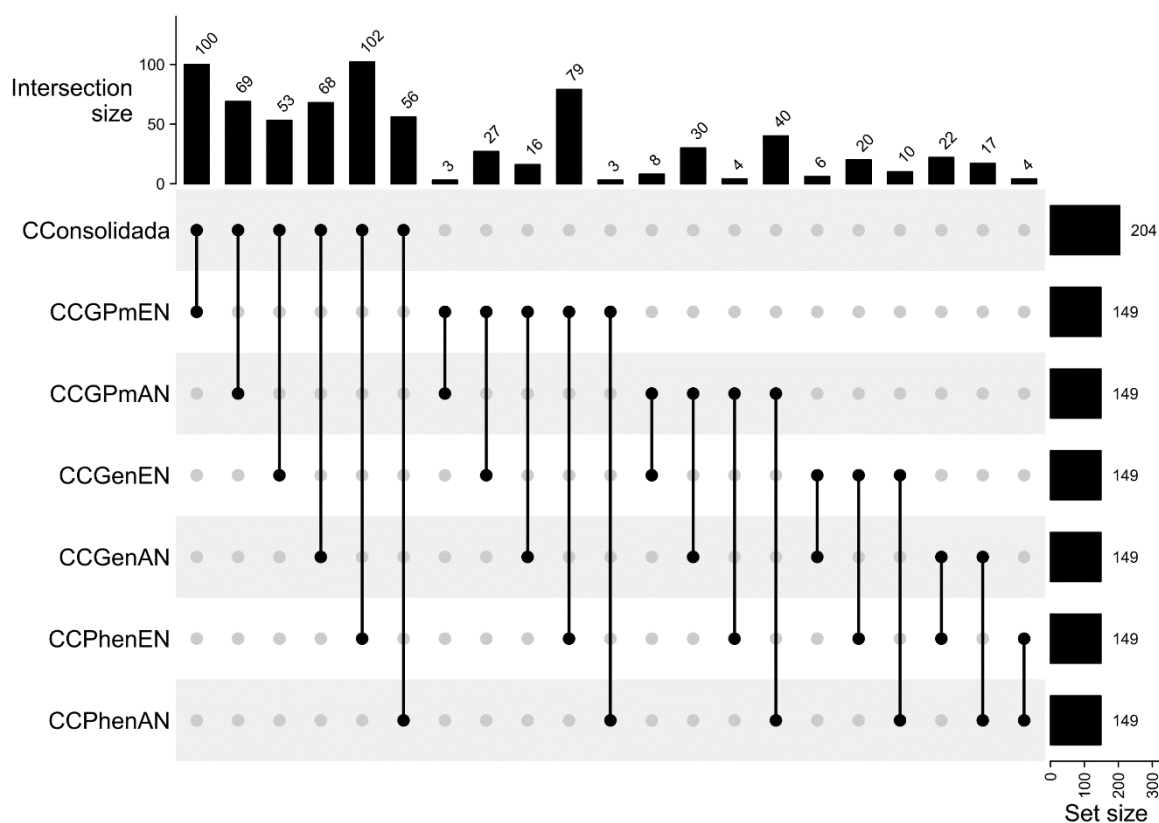
\*LA – latossolo amarelo; LV – latossolo vermelho; LVA – latossolo vermelho amarelo; DBC – Delineamento em blocos completos casualizados; DBA – Delineamento em blocos aumentados.



**Tabela S2.** Índice Kappa considerando diferentes metodologias para formação de coleções nucleares e coleção consolidada formada a partir dos acessos selecionados por pelo menos duas das abordagens utilizadas.

Índice Kappa	GenAN	GenEN	PhenAN	PhenEN	GPmAN	GPmEN	CCons
<b>GenAN</b>	1	-0.07	0.02	0.05	0.11	0.01	0.19
<b>GenEN</b>	-0.07	1	-0.04	0.04	-0.05	0.09	0.08
<b>PhenAN</b>	0.02	-0.04	1	-0.08	0.19	-0.09	0.10
<b>PhenEN</b>	0.05	0.04	-0.08	1	-0.08	0.48	0.44
<b>GPmAN</b>	0.11	-0.05	0.19	-0.08	1	-0.09	0.20
<b>GPmEN</b>	0.01	0.09	-0.09	0.48	-0.09	1	0.43
<b>CCons</b>	0.19	0.08	0.1	0.44	0.20	0.43	1

\*GenAN – Coleção nuclear formada por dados genotípicos e estratégia de otimização AN - *Average accession-to-nearest-entry*; GenEN – Coleção nuclear formada por dados genotípicos e estratégia de otimização EN - *Averageentry-to-nearest-entry*; PhenAN Coleção nuclear formada por dados fenotípicos e estratégia de otimização AN - *Average accession-to-nearest-entry*; PhenEN - Coleção nuclear formada por dados fenotípicos e estratégia de otimização AN - *Average accession-to-nearest-entry*; GPmAN – Coleção formada por dados morfoagronômicos +SNPs e estratégia de otimização AN - *Average accession-to-nearest-entry*; GPmEN - Coleção formada por dados morfoagronômicos +SNPs e estratégia de otimização EN - *Averageentry-to-nearest-entry*; CCons - Coleção consolidada que inclui acessos selecionados por pelo menos duas das abordagens utilizadas.



**Figura S1.** Acessos coincidentes considerando a coleção consolidada e as diferentes metodologias para formação de coleções nucleares: Coleções

nucleares obtidas exclusivamente com dados fenotípicos e algoritmo de otimização *Average entry-to-nearest-entry distance* (PhenEN) e *Average accession-to-nearest-entry distance* (PhenAN); exclusivamente com dados genotípicos (GenAN e GenEN); e coleções agrupadas com ambos dados fenotípicos e genotípicos (GPmAN e GPmEN).

## CAPÍTULO 2

**Desenvolvimento de coleções nucleares temáticas para atributos produtivos, resistência a doenças e qualidade de amido**

## Desenvolvimento de coleções nucleares temáticas para atributos produtivos, resistência a doenças e qualidade de amido em mandioca

### RESUMO

Coleções temáticas (CT) com tamanho reduzido e com os genótipos de maior interesse agrônomo podem otimizar a seleção de parentais nos programas de melhoramento de plantas. Para a utilização efetiva dos recursos de germoplasma de mandioca, três CT foram desenvolvidas, considerando os principais atributos de produtividade (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality). Os *best linear unbiased predictors* (BLUP) dos genótipos de mandioca foram ranqueados e feita a seleção truncada para cada coleção com base em características específicas. Em seguida, duplicatas foram eliminadas com base na análise de *identity by state* (IBS). A coincidência entre as CTs foi obtida pelo índice Kappa e a comparação entre as diferentes coleções com base na dispersão das características e no índice de diversidade de Shannon-Weaver (ISW). A diversidade genética foi avaliada por meio de parâmetros de heterozigosidade observada ( $H_o$ ); diversidade genética dentro da população ( $H_s$ ) e coeficiente de endogamia ( $F_{is}$ ). A estrutura das coleções foi avaliada na análise de componentes principais (PCA) e análise discriminante de componentes principais (ADCP). As coleções apresentaram baixa coincidência entre si, sendo cada uma delas composta por 72 (CC\_Disease), 63 (CC\_Root quality) e 64 acessos (CC-Yield), correspondendo a 4%, 3,5% e 3,5% dos indivíduos da coleção completa. Os valores de ISW variaram em sua maioria, mas não acima de 10% em comparação com a coleção completa. A maioria das CTs apresentaram parâmetros genéticos ( $H_o$ ,  $H_s$  e  $F_{is}$ ) muito semelhantes em comparação com a coleção completa, sendo capaz de reter 90.76%, 88.10% e 88.99% dos alelos da coleção completa (CC\_Disease, CC\_Root quality e CC\_Disease, respectivamente). As PCAs dos dados moleculares e agromorfológicos demonstraram boa distribuição e dispersão dos genótipos, enquanto, a ADCP possibilitou alta capacidade de discriminação dos acessos de cada coleção, possibilitando observar a separação entre as três coleções temáticas, sem grandes sobreposições. As coleções temáticas foram eficientes em reunir acessos com maior potencial produtivo, qualidade de raízes e resistência a doenças. Dessa forma, as estratégias adotadas neste trabalho têm grande potencial de uso no melhoramento da cultura.

**Palavras-chave:** *Manihot esculenta* Crantz, germoplasma, melhoramento, alto potencial de melhoramento.

## Development of thematic core collections for yield, disease resistance and starch quality traits in cassava

### ABSTRACT

Thematic collections (TC) with reduced size and comprising genotypes of higher agronomic interest can optimize parental selection in plant breeding programs. In order to effectively utilize cassava germplasm resources, three TCs were developed, focusing on the key attributes of yield (CC-Yield), pest and disease resistance (CC\_Disease), and root quality traits (CC\_Root quality). The cassava genotypes were ranked using the best linear unbiased predictors (BLUP) method, and a truncated selection was applied to each collection based on specific traits. Subsequently, duplicates were eliminated through identity by state (IBS) analysis. The degree of overlap between the TCs was assessed using the Kappa index, while comparisons among different collections were made based on trait dispersion and the Shannon-Weaver Diversity Index (ISW). Genetic diversity was evaluated through observed heterozygosity ( $H_o$ ), genetic diversity within the population ( $H_s$ ), and the inbreeding coefficient ( $F_{is}$ ). The structure of the collections was examined using Principal Component Analysis (PCA) and Discriminant Analysis of Principal Components (DAPC). The TCs exhibited limited overlap with each other, with each collection consisting of 72 (CC\_Disease), 63 (CC\_Root quality), and 64 genotypes (CC-Yield), which corresponded to 4%, 3.5%, and 3.5% of the total individuals in the complete collection, respectively. The ISW values mostly varied but did not exceed 10% in comparison to the complete collection. The majority of TCs demonstrated genetic parameters ( $H_o$ ,  $H_s$ , and  $F_{is}$ ) that closely resembled those of the complete collection, effectively retaining 90.76%, 88.10%, and 88.99% of the alleles present in the complete collection (CC\_Disease, CC\_Root quality, and CC\_Disease, respectively). PCA analysis of molecular and agro-morphological data indicated a well-distributed and dispersed distribution of genotypes, while DAPC exhibited a high discrimination capacity among the accessions within each collection, revealing clear separations between the three thematic collections with minimal overlap. The thematic collections proved to be efficient in gathering accessions with a greater potential for productivity, root quality, and disease resistance. Thus, the strategies employed in this study hold significant potential for crop improvement purposes.

**Keywords:** *Manihot esculenta* Crantz, germplasm, breeding, high breeding potential.

## 1.0 INTRODUÇÃO

A mandioca (*Manihot esculenta* Crantz) é uma das fontes de carboidratos mais importantes dos países tropicais, cuja produção tem crescido no mercado mundial por ser uma importante fonte de amido para indústrias de alimentos e inúmeras aplicações industriais não alimentares, além de ser um componente para ração animal e produção de bioetanol (CEBALLOS et al., 2020). A mandioca também é um alimento básico para cerca de 800 milhões de pessoas, sobretudo na África, América Latina e Ásia (ADEYEMO et al., 2017).

A produção brasileira de mandioca em 2022 foi 18,49 milhões de toneladas, colhidas em uma área de 1,24 milhão de hectares (IBGE, 2022). Com o aumento da inserção da cultura no agronegócio brasileiro, cresceu-se a preocupação com o uso de variedades melhoradas adaptadas às mais diversas regiões de cultivo no país, bem como a disponibilidade de material propagativo em quantidade e no momento adequado. O trabalho de pesquisa tem avançado bastante e nas últimas décadas a disponibilidade de variedades melhoradas de mandioca no Brasil aumentou consideravelmente (VIEIRA et al., 2019; OLIVEIRA et al., 2020).

Os ganhos genéticos no melhoramento de mandioca dependem da exploração eficiente da variabilidade genética disponível nos bancos de germoplasma (BAG). O acesso à variabilidade genética armazenada nos BAGs é fundamental para o melhoramento da cultura, principalmente para seleção de clones com alto breeding value e melhor desempenho agrônômico *per se*. De acordo com Boukar et al. (2013) a crescente demanda por variedades melhoradas cada vez mais produtivas, com resistência às principais pragas e doenças, adequação a sistemas de cultivo e manejo mais rentáveis, têm pressionado os bancos de germoplasma a serem cada vez mais eficientes, sobretudo, nas caracterizações e avaliações do germoplasma para entrega de parentais com alelos associados a novos atributos ou mesmo alelos que possam melhorar características produtivas e/ou de agregar maior valor econômico. Por outro lado, extensivas caracterizações e avaliações, em grandes coleções de germoplasma de mandioca são inviáveis devido ao custo e tempo necessário para obtenção de dados fenotípicos confiáveis. O uso de coleções de menor tamanho e focada em atributos específicos parece ser uma estratégia bastante

adequada para otimizar e concentrar os esforços de caracterização, avaliação e uso em acessos de maior valor.

O conceito de coleção nuclear foi introduzido por Frankel e Brown (1984), como um subconjunto mínimo de indivíduos que represente a variabilidade máxima de toda coleção de germoplasma. O conceito tem objetivo de contornar e facilitar o manuseio do grande número de acessos mantidos em grandes coleções, que tornam os processos de conservação, catalogação, caracterização, e avaliação, extremamente laboriosos e de alto custo. Com a diversidade genética e morfológica de toda coleção representada em uma pequena amostra de acessos, cuidadosamente avaliados e descritos, é possível aumentar a introdução e o uso destes recursos genéticos nos programas de melhoramento. Entretanto, com o tempo e a introdução de novos acessos, as coleções nucleares tendem a ficar maiores do que o planejado e assim perpetuar o problema de grandes coleções com alto custo de manutenção e baixo uso na geração de produtos e tecnologias (PESSOA FILHO et al., 2010). Por isso, coleções nucleares temáticas têm sido propostas e visam aumentar a eficiência na identificação de diversidade para características específicas e prioritárias nos programas de melhoramento (BOCZKOWSKA et al., 2016).

De acordo com Pessoa-Filho et al. (2010) grandes coleções de germoplasma podem ter várias coleções nucleares temáticas de tamanhos menores, com foco em diferentes características de interesse. Atualmente, diversos programas de melhoramento genético vêm adotando essa estratégia em outras espécies como: *Olea europaea* L. (EL BAKKALI et al., 2013); soja (GUO et al., 2014); sorgo (UPADHYAYA et al., 2019); e grão-de-bico (UPADHYAYA et al., 2006). Para mandioca a estratégia traria benefícios importantes para os programas de melhoramento genético, promovendo o uso eficiente dos recursos disponíveis, amenizando a presença de acessos redundantes, resultantes do alto intercâmbio de germoplasma entre agricultores e frequente troca de nomes para um mesmo acesso. Recentemente, Perez-Fons et al. (2020), desenvolveram e demonstraram o potencial de uma coleção básica de germoplasma utilizando informações genômicas e metabolômicas como ferramenta de fenotipagem, para caracterizar materiais elite e de pré-melhoramento, com alto potencial de uso destes acessos como parentais em programas de melhoramento direcionados ao aumento da tolerância a estresses

bióticos e abióticos.

A utilização eficiente das variações genéticas disponíveis nos bancos de germoplasma podem ser exploradas permitindo aos melhoristas enfrentar os desafios e cenários futuros associados à maior vulnerabilidade genética a estresses bióticos e abióticos (WANG et al., 2017), além de ajudar no desenvolvimento de perfis de produto voltados para atendimento das demandas dos usuários finais das variedades de mandioca, bem como aos cenários de expansão e crescimento da espécie. Assim, o objetivo deste trabalho foi: i) desenvolver coleções nucleares temáticas que possam armazenar a máxima diversidade possível para atributos produtivos, resistência a doenças e qualidade das raízes, e ii) comparar a diversidade dessas coleções temáticas com a coleção completa.

## **2.0 MATERIAL E MÉTODOS**

### **2.1 Dados morfo-agronômicos e avaliação do material vegetal**

O estudo envolveu uma coleção com 1.665 acessos de mandioca pertencentes ao Banco Ativo de Germoplasma e programa de melhoramento da Embrapa Mandioca e Fruticultura, localizada na cidade de Cruz das Almas, Bahia (Latitude: 12°40'19"S, Longitude: 39°06'22"W e Altitude: 220 m). Os descritores avaliados para a formação das coleções nucleares temáticas foram escolhidos de acordo com sua associação com atributos produtivos, resistência a doenças e qualidade das raízes de acordo com Fukuda et al. (2010).

Atributos associados à produtividade dos genótipos:

1. AP – Altura da planta, avaliado do solo até o meristema da planta
2. PI.Por – porte da planta, avaliado por meio de uma escala de notas (1-5), sendo 1 = porte excelente (sem ramificações ou com ramificações acima de 2.0 m), 2 = porte bom (ramificações acima de 1.60 m ou ramificação baixa, mas com pelo menos 1.6 m de hastes eretas), 3 = porte mediano (ramificações acima de 1.20 m ou ramificação baixa, mas com pelo menos 1.2 m de hastes eretas), 4 = porte ruim (ramificações acima de 0.80 m ou ramificação baixa, mas com pelo menos 0.80 m de hastes eretas), e 5 =



- porte muito ruim (clones muito ramificados com menos de 0.80 m de hastes eretas)
3. PI.Ty – tipo de planta, , avaliado por meio de uma escala de notas (1-5), sendo 1 = compacta, 2 = aberta, 3 = guarda-sol, 4 = cilíndrica
  4. PPA – produtividade de parte aérea ( $t.ha^{-1}$ ) mensurada como o peso total de parte aérea, incluindo folhas, pecíolos e hastes
  5. Ro.PRC – produtividade de raízes comerciais ( $t.ha^{-1}$ ), caracterizada pela ausência de sintomas de pragas e doenças, sem danos mecânicos oriundos da colheita, tamanho e formato padrão do genótipo
  6. Ro.PRNC – produtividade de raízes não comerciais ( $t.ha^{-1}$ )
  7. PTR – produtividade total de raízes ( $t.ha^{-1}$ ) = Ro.PRC + Ro.PRNC
  8. TMS – teor de matéria seca nas raízes (%), mensurada pelo método da balança hidrostática (KAWANO et al., 1987)
  9. StC – teor de amido (%), mensurada pelo método da extração de amido das raízes (VASCONCELOS et al., 2017)
  10. PMS – produtividade de matéria seca ( $t.ha^{-1}$ ), mensurado pelo produto da Ro.PTR x Ro.DMC
  11. IC – índice de colheita que é a relação entre o peso de raízes e o peso total da planta,  $IC = \frac{Pesototalderaízes}{Pesototaldaplanta}$
  12. ECR – espessura da entrecasca (córtex) das raízes (mm)
  13. PPR – presença de pedúnculo nas raízes, avaliado por meio de uma escala de notas, em que 0 = sésil, 3 = pedunculada, 5 = mistura de raízes sesséis e pedunculadas
  14. CR – Constrição das raízes, avaliado por meio de uma escala de notas (1-3), sendo 1 = poucas ou nenhuma constricção, 2 = algumas constricções, 3 = muitas constricções
  15. CMR – comprimento médio das raízes (cm), mensurada na base até a ponta das raízes
  16. DMR – diâmetro médio das raízes (mm), mensurado na parte central das raízes
  17. DPR – destaque da casca das raízes avaliada na época da colheita, por escala de notas, em que 3 = descascamento fácil, 5 = descascamento médio, 7 = descascamento difícil

18. PI.RF – retenção foliar com uso de escala de notas, em que 1 = menos 5% de retenção foliar, 2 = entre 6 e 15% de retenção foliar, 3 = entre 16 e 30% de retenção foliar, 4 = entre 31 e 50% de retenção foliar, 5 = mais de 50% de retenção foliar
19. PI.Vigor1.5M – vigor das plantas avaliado aos 1,5 meses após o plantio com base em escala de notas em que 1 = baixo vigor, 3 = vigor intermediário, 5 = alto vigor
20. PL.Vigor12M – vigor das plantas avaliado aos 12 meses após o plantio com base na escala de notas da característica PI.Vigor1.5M
21. NR - número médio de raízes por planta

**Atributos associados à resistência a pragas e doenças:**

1. D.AcSev – avaliação da severidade causada por ácaros, avaliado de acordo com a seguinte escala de notas, 0 = sem danos, 1 = algumas pontuações branco-amarelas até a base das folhas do broto apical, 2 = pontuações amarelas moderadas em todas as folhas, 3 = pontuações abundantes nas folhas do terço médio, ligeira deformação do broto apical, 4 = severa deformação das folhas do broto apical, folhas com aparência esbranquiçada e alguma desfolha, e caule com pontuações amarelas, 5 = brotos apicais muito retorcidos em forma de repolho/severas pontuações no caule, 6 = brotos apicais secos, sem folha ou mortos
2. D.Anth – severidade causada pela antracnose (*Colletotrichum* sp.) de acordo com a escala de notas: 1 = sem sintoma, 2 = cancrs pequenos ou antigos na metade inferior da planta, 3 = cancrs profundos na metade superior da planta, 4 = cancrs profundos, com esporulação, distorção ou murcha nas folhas novas e secamento do ápice, 5 = morte apical ou morte da planta
3. D.BacSev – severidade causada pela bacteriose (*Xanthomonas axonopodis* pv. *manihotis*) de acordo com a escala de notas: 1 = sem sintomas, 2 = presença apenas de sintomas foliares (mancha foliar angular), 3 = presença de lesões necróticas escuras em hastes ou pecíolos, 4 = sintomas graves nas folhas e/ou presença de lesões necróticas com exsudação de goma, 5 = perda total das folhas com morte apical ou morte da planta

- D.MP – severidade da mancha parda (*Passalora henningsii*), de acordo com a escala de notas: 0 = sem sintoma 1 = sintomas brandos no terço inferior 2 = manchas foliares no terço inferior e amarelecimento em poucas folhas afetadas 3 = manchas foliares nos terços inferior e médio e amarelecimento na maioria das folhas afetadas 4 = manchas foliares distribuídas por toda a planta, com amarelecimento e queda das folhas do terço inferior 5 = desfolha completa da planta
- 4. D.QF – severidade da mancha parda (*Passalora viçosae*), avaliada com a mesma escala da D.MP
- 5. D.FerSev – severidade da ferrugem (*Uromyces manihotis*), de acordo com escala de notas, em que 0 = ausência de pústulas e danas nas hastes e folhas, 1 = presença de poucas pústulas, sem seca de ramos, 2 = presença de muitas pústulas, com seca de ramos
- 6. D.FrogSkin – severidade causada pelo couro de sapo (complexo de vírus e fitoplasmas), com base na escala de notas, em que 0 = ausência raízes com sintomas, 1 = presença de poucas raízes com epiderme corticosa, 2 = presença de muitas raízes com epiderme corticosa, com profundas rachaduras longitudinais ou ausência na produção de raízes tuberosas
- 7. D.Pod – severidade causada pela podridão mole (*Phytophthora* sp. *Pythium* sp. e *Phytopythium* sp), seca (*Fusarium* sp) e negra das raízes (*Neoscytalidium hyalinum* e *Lasiodiplodia* sp.), de acordo com a escala de notas, 0 = ausência de podridão radicular, 1 = até 25% das raízes com podridão radicular, 2 = acima de 26% das raízes com podridão radicular

#### **Atributos associados à qualidade das raízes:**

1. Ro.Cook – tempo de cozimento das raízes avaliados aos 15, 20, 25, 30 e 40 de acordo com Talma et al. (2013)
2. CPR – cor da polpa das raízes (1 = branca, 2 = creme, 3 = amarela, 4 = rosada, 5 = alaranjada)
3. CER – cor da casca das raízes (1 = Branco 2 = Amarelo 3 = Marrom claro 4 = Marrom escuro)
4. CCR – cor do córtex das raízes (1 = branco ou creme 2 = amarelo 3 – rosado 4 – roxo)
5. Ro.Friab – avaliação da friabilidade das raízes cozidas (FRANCK et al.,

- 2011, com base em escala de notas, em que 1 = muita coesa, sem rachaduras, 2 = raízes coesas com centro rachado, 3 = friável, centro rachado e a superfície das raízes parcialmente rachada, 4 = muito friável, maior parte do centro e superfície das raízes está rachada
6. HCN – teor de ácido cianídrico nas raízes, determinado pelo método rápido do picrato, conforme Bradbury et al. (1999)
  7. Ro.TCC – teor de carotenoides totais nas raízes de acordo com Carvalho et al. (2022a).

## 2.2 Análise de dados

Como os acessos foram caracterizados em diferentes anos de cultivo (2011 a 2021- Tabela S1), algumas características qualitativas apresentaram ligeira variação ao longo das avaliações e por isso, utilizou-se a moda dos dados para caracterizar determinados acessos de mandioca. Por outro lado, as características quantitativas foram analisadas por modelos lineares mistos, devido ao grau de desbalanceamento presente nos ensaios de avaliação do germoplasma de mandioca.

O germoplasma de mandioca foi analisado em diversos ambientes em diferentes anos de cultivo, por isso a combinação de local versus ano de cultivo foi chamada de “ambiente” daqui em diante. A análise conjunta considerando todos os ambientes para cada descritor quantitativo foi realizada com base no modelo:  $y = Zg + Wb + Ti + e$ , no qual:  $y$  é o vetor de observações fenotípicas;  $g$  é o vetor de efeitos genotípicos, considerado como aleatório  $gN(0, \sigma_g^2)$ ;  $b$  é o vetor de efeitos alinhados de blocos dentro de ensaios, considerado como aleatório  $bN(0, \sigma_b^2)$ ;  $i$  é o vetor de efeitos da interação genótipos x ensaios, considerado como aleatório  $(0, \sigma_{ge}^2)$ ;  $e$  é o vetor de efeitos de erros, considerado como aleatório  $eN(0, \sigma_e^2)$ .  $Z$ ,  $W$ ,  $T$  são as matrizes de incidência para os referidos efeitos, respectivamente. Esse modelo foi utilizado para obter as estimativas dos valores genéticos dos genótipos oriundos da avaliação de experimentos sob delineamento de blocos incompletos utilizados em vários ensaios. As análises de modelo linear misto foram realizadas por meio do pacote *sommer* do software R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022).

### 2.3 Definição das coleções temáticas

Foram formadas três coleções temáticas, considerando os principais atributos de produtividade (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality). A coleção CC-Yield foi construída com base na seleção truncada dos 12 clones com menores BLUPs para porte das plantas e maiores BLUPs para PTR, PMS, PL.Vigor12M, PI.Vigor1.5M, TMS e NR. A coleção CC\_Disease também considerou a seleção truncada dos genótipos com menores BLUPs para severidade causada por ácaros (D.AcSev) e menor severidade para doenças da parte aérea como D.Anth, D.BacSev, D.MP, D.QF e D.FerSev, além de doenças radiculares como D.FrogSkin. Para construção da CC\_Root quality, foi levada em consideração a seleção de genótipos com menores BLUPs para HCN, Ro.Cook abaixo de 25 min. Em seguida, os genótipos selecionados anteriormente sofreram uma nova seleção truncada para seleção dos 12 genótipos com CPR creme escuro e amarela, com maior Ro.TCC e TMS, além de raízes friáveis ou muito friáveis após cozimento.

### 2.4 Análise de parentesco entre acessos e coincidência entre as coleções

Os dados fenotípicos coletados para formação das coleções temáticas, bem como marcadores do tipo *single nucleotide polymorphism* (SNP), reportados por Ogbonna et al. (2021) foram utilizados agrupamento das coleções e análise de parâmetros genéticos. As análises das sequências e filtros de qualidade dos dados SNPs foram realizadas no software Tassel versão 5.2.37 (BRADBURY et al., 2007) com remoção dos alelos de frequência mínima (MAF) e dados perdidos (Call Rate) inferior a 5% e superior a 80%, respectivamente. Posteriormente, os dados perdidos foram imputados no software Beagle 4.1 (BROWNING, 2016). Um total de 20.023 SNPs distribuídos nos 18 cromossomos da mandioca foram utilizados nas análises.

Após a formação das coleções, o grau de parentesco entre os acessos selecionados foi avaliado por IBS (*Identity-by-state*) indivíduo a indivíduo usando a função *snpGdsIBS* no pacote *SNPRelate* (ZHENG et al., 2012). Os pares de acessos com elevado grau de parentesco (>0.95) foram avaliados e um dos

indivíduos retirados das análises posteriores para que fosse possível obter coleções temáticas com o mínimo de redundância dos acessos.

A coincidência dos genótipos entre as coleções foi realizada usando um código binário, na qual os indivíduos selecionados e não selecionados receberam os códigos 1 e 0, respectivamente. A coincidência entre as três coleções temáticas foi obtida pelo índice Kappa (COHEN, 1960). Todas as análises foram realizadas no R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022).

## 2.5 Parâmetros genéticos das coleções temáticas

A comparação entre as diferentes coleções temáticas e a coleção completa para os dados fenotípicos foi realizada com base na dispersão das características quantitativas e qualitativas, e com base nos índices de diversidade de Shannon-Weaver para cada característica nas coleções completa e coleções individuais, de acordo com a seguinte fórmula  $H' = -\sum_{i=1}^n p_i \log_e p_i$ : em que  $p_i$  é a frequência observada da classe  $i$  para determinada característica,  $n$  é o número das classes fenotípicas da característica,  $p_i$  é a frequência relativa de cada descritor. Todos os índices  $H'$  foram normalizados e divididos pelo valor máximo visando manter os valores entre 0 e 1 (monomorfismo até máxima diversidade fenotípica). Para as características qualitativas,  $k$  representa o número de classes ou notas do descritor, já para os caracteres quantitativos foram estimadas seis classes com os limites inferior e superior definidos pela amplitude total observada na coleção completa para cada característica (Tabela 1). Todas as análises foram realizadas no R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022).

A diversidade genética nas coleções temáticas e coleção completa foi avaliada por meio dos 20.023 SNPs, utilizando os parâmetros de heterozigosidade observada ( $H_o = 1 - \sum k \sum i P_{kii}/np$ ); diversidade genética dentro da população  $H_s = \bar{n}/(\bar{n} - 1) \left[ 1 - \sum_i P_i^{-2} - \frac{H_o}{2\bar{n}} \right]$ ; e coeficiente de endogamia ( $F_{is} = 1 - H_o/H_s$ ), implementados com o auxílio do pacote *hierfstat* (GOUDET, 2005) no R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022).

**Tabela 1.** Estimação dos limites inferiores e superiores das classes fenotípicas a serem utilizadas para a estimação do índice de diversidade de Shannon-Weaver para as características quantitativas.

Classe fenotípica	Limite Inferior	Limite Superior
1	$min$	$\leq min + Amp/6$
2	$min + Amp/6$	$\leq min + 2x Amp/6$
3	$min + 2x Amp/6$	$\leq min + 3x Amp/6$
4	$min + 3x Amp/6$	$\leq min + 4x Amp/6$
5	$min + 4x Amp/6$	$\leq min + 5x Amp/6$
6	$min + 5x Amp/6$	$\leq max$

min: mínimo da característica da coleção completa, max: máximo da característica da coleção completa, Amp: amplitude da característica da coleção completa.

## 2.6 Diversidade fenotípica e molecular das coleções temáticas de mandioca

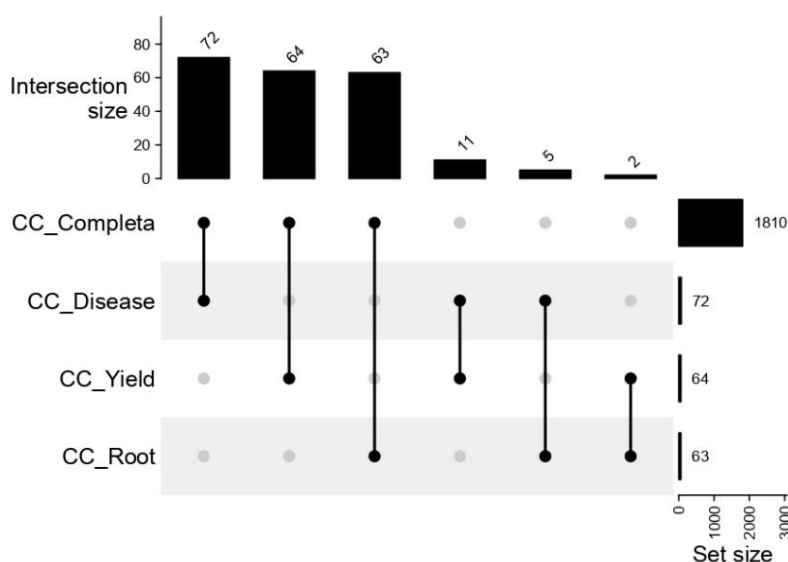
A estrutura das coleções temáticas e coleção completa foi avaliada com base na análise de componentes principais (PCA) dos dados fenotípicos realizada via pacote *AMR* (BERENDS et al., 2022), e dos dados moleculares implementada com auxílio do pacote *PCAtools* (BLIGHE, 2020) ambos no software R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022).

Adicionalmente a análise discriminante de componentes principais (ADCP) disponível no pacote *adegenet* para R versão 4.2.2 (R DEVELOPMENT CORE TEAM, 2022) foi utilizada para discriminar as coleções temáticas de mandioca e a coleção completa, considerando os dados fenotípicos quantitativos e moleculares. Foram retidos os eixos da análise de componentes principais que explicaram mais de 80% da variância total dos dados e os genótipos de mandioca foram designados a cada uma das três coleções temáticas (CC-Yield, CC\_Disease e CC\_Root quality) como agrupamento *a priori*.

### 3.0 RESULTADOS

#### 3.1 Formação das coleções temáticas

Com intuito de utilizar os recursos de germoplasma de mandioca de forma eficiente, foram desenvolvidas três coleções temáticas levando em consideração os principais atributos de produtividade (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality) selecionando inicialmente 70, 79 e 79 acessos, respectivamente. Em seguida, os acessos selecionados foram avaliados quanto à sua redundância de acordo com a IBS, com o objetivo de eliminar duplicatas dentre os indivíduos selecionados em cada coleção temática. A análise molecular revelou alto grau de parentesco ( $>0.95$ ) entre 108 pares de acessos (Tabela S2) o que resultou na exclusão de 29 acessos de mandioca. Assim, 199 acessos foram selecionados pelas três coleções temáticas, sendo 72 (CC\_Disease), 63 (CC\_Root quality) e 64 acessos (CC-Yield), correspondendo a 4%, 3,5% e 3,5% dos indivíduos com informações fenotípicas da coleção completa (1.810 acessos), respectivamente (Figura 1).



**Figura 1.** Número de acessos coincidentes considerando diferentes coleções temáticas: CC\_Disease – Coleção temática para resistência a pragas e doenças; CC\_Root\_quality – Coleção temática para qualidade de raízes; CC\_Yield – Coleção temática para atributos de produtividade.

No caso dos dados moleculares, 347 acessos do germoplasma de mandioca não foram genotipados e com isso apenas 164 dos 1.463 indivíduos



genotipados foram utilizados para avaliação dos parâmetros genéticos nas coleções temáticas, sendo 60, 56 e 48 acessos nas coleções CC\_Disease, CC\_Yield e CC\_Root quality, respectivamente.

Por se tratar de características não diretamente correlacionadas, houve baixa coincidência na seleção dos acessos para compor as diferentes coleções temáticas, que foram formadas por acessos exclusivos, na maioria dos casos. Essa observação foi corroborada pelo índice Kappa que variou de 0.16 a 0.048 (Tabela 2). As maiores concordâncias na seleção dos acessos de mandioca foram entre as coleções temáticas CC\_Disease e CC\_Yield (0.176) tendo 11 acessos em comum. Já as coincidências entre as demais coleções foram 0.096 entre as coleções CC\_Root\_quality e CC\_Disease (5 acessos) e 0.11 entre as coleções CC\_Root\_quality e CC\_Yield (2 acessos) (Figura 1). A baixa coincidência de acessos entre as coleções temáticas, reforçam, portanto, que a formação de coleções temáticas com características específicas desejáveis é determinante para o sucesso do melhoramento da cultura.

**Tabela 2.** Índice Kappa de coincidência entre acessos selecionados por diferentes coleções temáticas de mandioca.

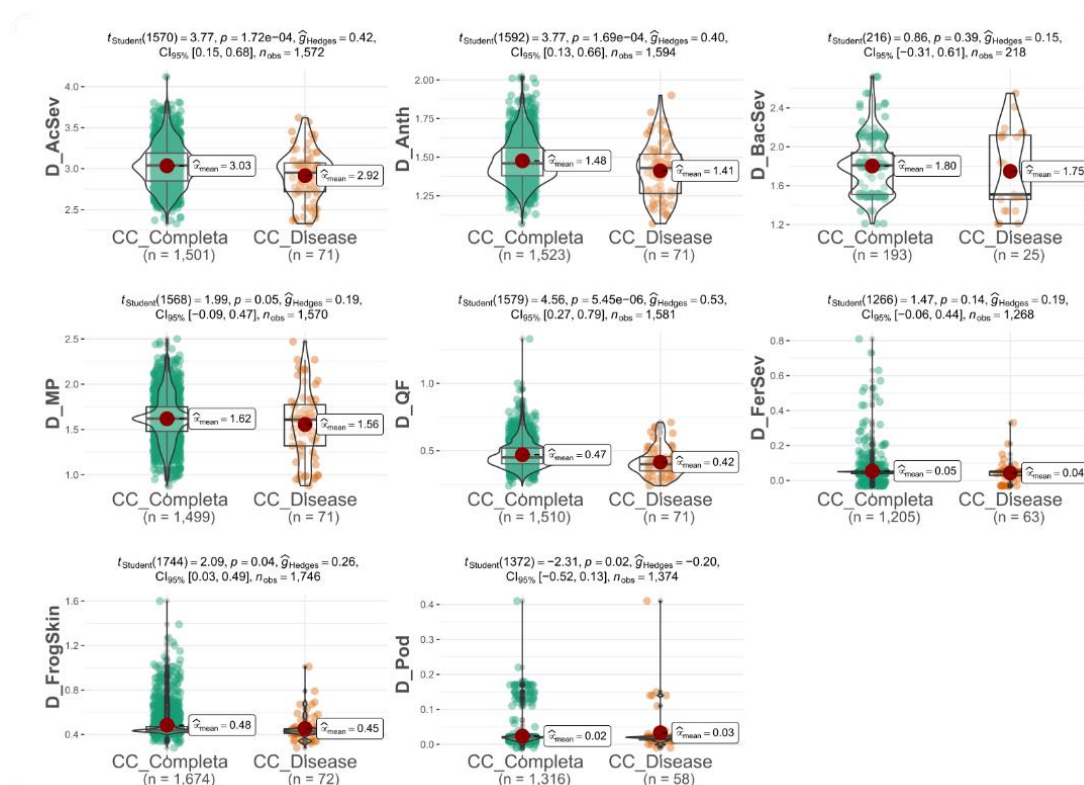
Índice Kappa	CC_Disease	CC_Root_quality	CC_Yield
CC_Disease	1	0.096	0.176
CC_Root_quality	0.096	1	0.110
CC_Yield	0.176	0.110	1

CC\_Disease – Coleção temática para resistência a doenças e pragas; CC\_Root\_quality – Coleção temática para qualidade de raízes; CC\_Yield – Coleção temática para atributos produtivos.

### 3.2 Variação fenotípica nas coleções temáticas

Para desenvolvimento da coleção temática para resistência a doenças e pragas (CC\_Disease), foram selecionados apenas indivíduos que apresentaram menores sintomas das doenças avaliadas. Como esperado, para maioria das doenças houve um deslocamento da curva em direção à redução da média e menores valores de severidade de pragas e doenças em comparação a coleção completa, principalmente para a severidade causada por ácaros, antracnose, bacteriose, mancha parda e queima das folhas (Figura 2). De fato, o escore médio da severidade causada por ácaros foi de 2,92, enquanto a média geral da

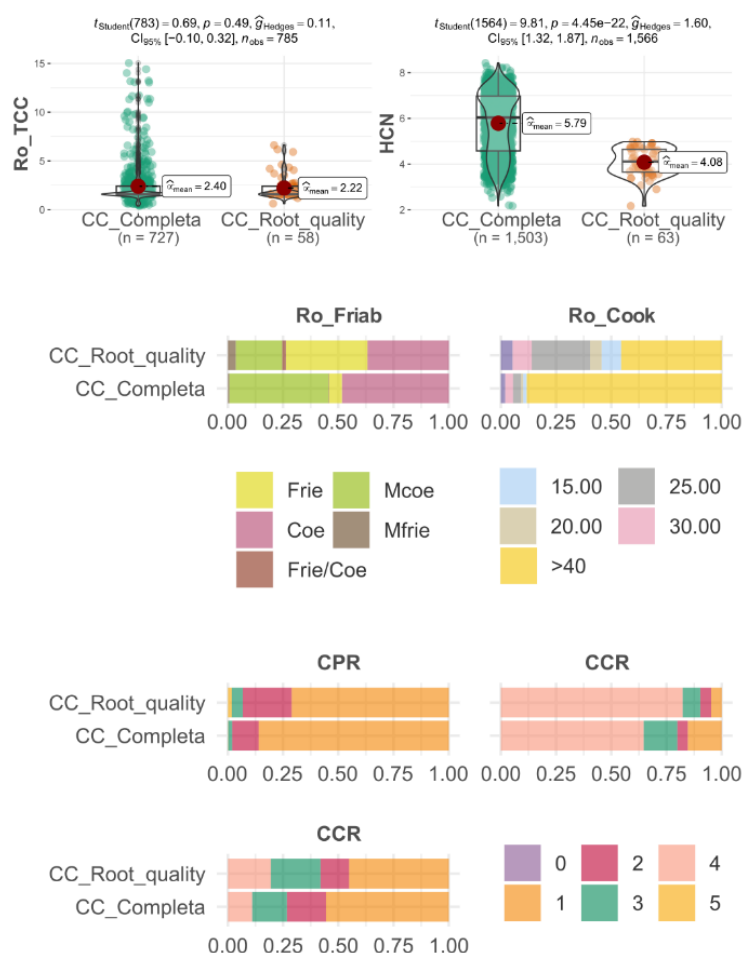
população foi de 3,03. Para as doenças da parte aérea, os escores médios nas coleções temáticas também foram ligeiramente menores do que as médias gerais da coleção completa: a antracnose apresentou score médio de 1,41 versus 1,48 da coleção completa; a bacteriose de 1,75 versus 1,80; a mancha parda de 1,56 versus 1,62; e a queima foliar de 0,42 versus 0,47.



**Figura 2.** Análise comparativa para diferentes atributos associados à resistência a pragas e doenças da mandioca, considerando a coleção temática para resistência a pragas e doenças (CC\_Disease) em comparação com a coleção completa (CC\_Completa). D\_AcSev: ácaros; D\_Anth: antracnose; D\_BacSev: bacteriose; D\_MP: mancha parda; D\_QF: mancha parda; D\_FerSev: ferrugem; D\_FrogSkin: couro de sapo; D\_Pod: podridão radicular.

As características relacionadas a coleção temática CC\_Root\_quality, foram selecionados em sua maioria com foco na conservação de genótipos de mesa que possuem melhores qualidades culinárias (Figura 3). Com isso, houve um deslocamento da curva no sentido da redução do teor de ácido cianídrico nas raízes, sendo selecionados basicamente clones com escores inferiores a 5 no teste do picrato que na prática representa acessos seguros para consumo humano. O score médio geral de teor de ácido cianídrico (HCN) foi de 5,79, enquanto a coleção temática apresentou média 4,08. Já para teor de carotenoides totais nas raízes (RO\_TCC), não houve alteração expressiva na

média da coleção temática CC\_Root\_quality (2,22) em comparação com a coleção completa (2,40). Acessos com os maiores valores de teor de carotenoides não foram selecionados nessa coleção, possivelmente por apresentarem alto teor de compostos cianogênicos e com isso não serem caracterizados como acessos para consumo fresco.



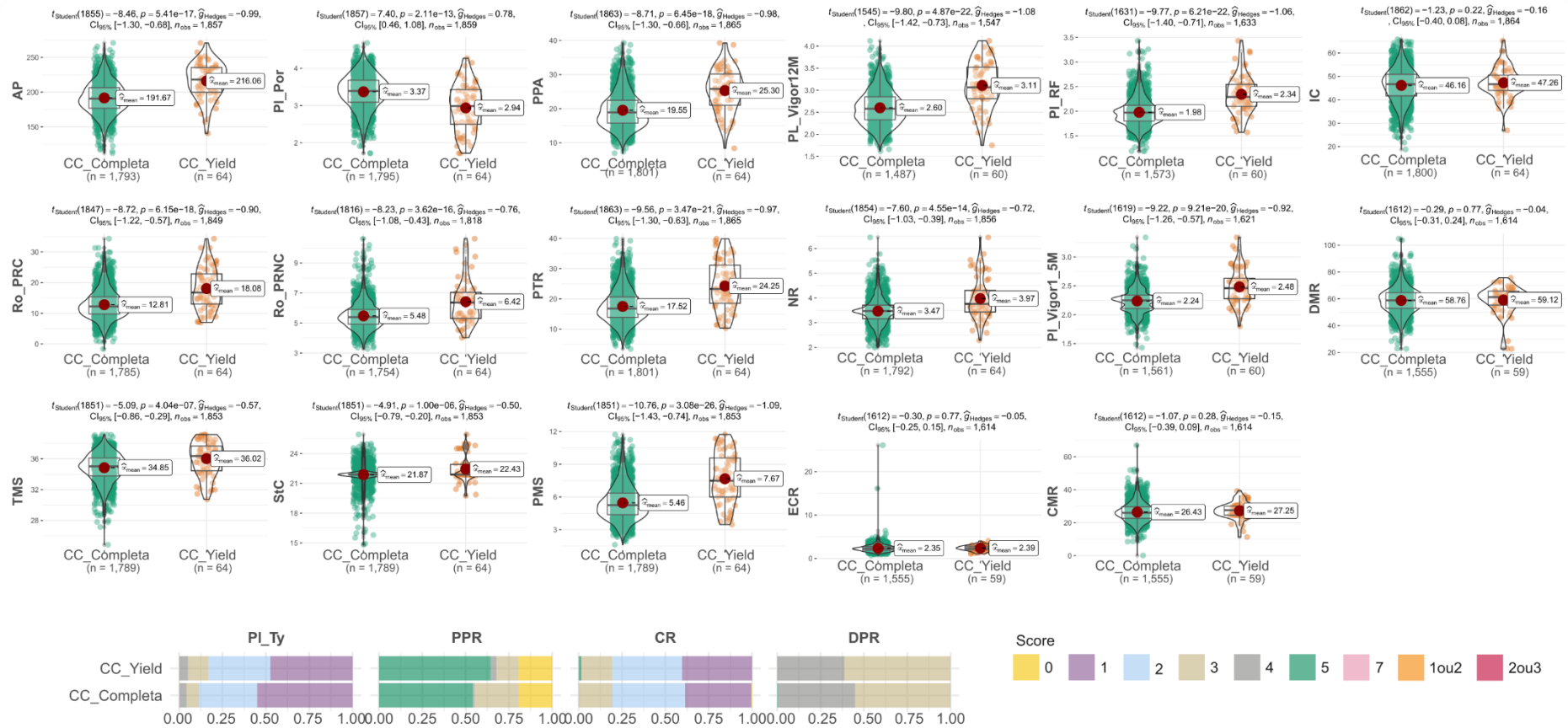
**Figura 3.** Análise comparativa para diferentes atributos associados à qualidade de raízes de mandioca, considerando a coleção temática qualidade de raízes (CC\_Root\_quality) em comparação com a coleção completa. HCN: teor de ácido cianídrico nas raízes; Ro\_TCC: teor de carotenoides totais nas raízes; Ro\_Friab avaliação da friabilidade das raízes cozidas; Ro\_Cook: tempo de cozimento das raízes; CPR: cor da polpa das raízes; CER: cor da casca das raízes; CCR: cor do córtex das raízes.

Em relação às demais características qualitativas da coleção temática CC\_Root\_quality, aumentou-se a frequência de clones com tempo de cozimento inferior a 30 minutos que é considerado o padrão brasileiro ideal para seleção de clones de mesa superiores. Como consequência, houve aumento na frequência

de clones com polpa considerada friável e muito friável que estão diretamente relacionados à elevada capacidade de cozimento. Em relação à cor da polpa das raízes, embora não tenha sido observada diferença na média do teor de carotenoides totais, a frequência de clones com polpa creme e amarela aumentou na coleção temática CC\_Root\_quality.

A coleção temática para atributos produtivos (CC-Yield) englobou acessos que possuem características importantes para o mercado de mandioca industrial. Semelhantemente às características das outras coleções temáticas, houve um deslocamento das curvas de distribuição das características em favor dos atributos produtivos, em especial àquelas que possuem uma alta variação na coleção completa, a exemplo da altura da planta (216,06 cm e 191,67 cm, na coleção temática e completa, respectivamente); produtividade de parte aérea (25,30 t.ha<sup>-1</sup> versus 19,55 t.ha<sup>-1</sup>, respectivamente); produtividade de raízes comerciais (12,8 versus 18,08 t.ha<sup>-1</sup>, respectivamente); produtividade de raízes não comerciais (6,42 versus 5,48 t.ha<sup>-1</sup>, respectivamente); produtividade de raízes frescas (24,25 versus 17,52 t.ha<sup>-1</sup>, respectivamente); produtividade de matéria seca (7,67 versus 5,46 t.ha<sup>-1</sup>, respectivamente); retenção foliar (2,34 versus 1,98, respectivamente); e teor de matéria seca nas raízes (36,02% versus 34,85%, respectivamente). O vigor das plantas da coleção CC-Yield avaliado aos 1,5 e 12 meses após o plantio também foi mais elevado na coleção temática em relação à coleção completa. Em contrapartida, os boxplots das características com menor variação entre os acessos de mandioca nas coleções temáticas em relação a coleção completa foram identificadas para espessura da entrecasca, índice de colheita e comprimento médio das raízes (Figura 4).

Para as características qualitativas, os acessos selecionados na coleção CC-Yield apresentaram apenas pequenas variações na proporção das classes. Na maioria dos casos, os acessos apresentaram fácil descascamento da película das raízes, raízes com poucas constrições e sem pedúnculo (séssil) e plantas tipo compacta ou aberta (Figura 4). No geral, esses resultados indicaram que os acessos de mandioca selecionados pelas coleções temáticas apresentam características especialmente de interesse agrônomo no melhoramento da espécie. Além de potenciais aplicações comerciais imediatas, essa seleção deve facilitar a conservação de acessos promissores para objetivo de cada coleção além de possibilitar uma caracterização fenotípica e genotípica mais exaustiva.



**Figura 4.** Análise comparativa para diferentes atributos agrônômicos e de produtividade de genótipos de mandioca, considerando a coleção temática atributos produtivos (CC-Yield) em comparação com a coleção completa. AP: altura da planta; PI\_Por: porte da planta; PPA: produtividade de parte aérea; Ro\_PRC: produtividade de raízes comerciais; Ro\_PRNC: produtividade de raízes não comerciais; PTR: produtividade de raízes frescas; PMS: produtividade de matéria seca; StC: teor de amido; IC: índice de colheita; Ro\_ECR: espessura da entrecasca; CMR: comprimento médio das raízes; DMR: diâmetro médio das raízes; PI\_RF: retenção foliar; PL\_Vigor1.5M e PL\_Vigor12M vigor das plantas avaliado aos 1,5 e 12 meses, respectivamente; TMS: teor de matéria seca nas raízes; NR: número médio de raízes; PI\_Ty: tipo de planta; PPR: presença de pedúnculo nas raízes; CR: constrição das raízes; DPR: destaque da casca das raízes.

### 3.3 Análise da diversidade fenotípica das coleções temáticas

As coleções temáticas desenvolvidas foram avaliadas quanto à diversidade genética mensurada pelo índice de Shannon-Weaver (ISW) (Tabela 3). Em sua maioria, não houve diferença superior a 10% nos valores do ISW nas coleções temáticas em comparação com a coleção completa, à exceção da característica ECR (espessura da entrecasca) na coleção temática CC\_Root\_quality, em que o ISW mais que dobrou. Por outro lado, o ISW de ECR foi nulo nas coleções CC\_Yield e CC\_Disease, indicando a ausência de variância para esta características nos clones de mandioca selecionados por estas coleções temáticas.

**Tabela 3.** Índices de Shannon-Weaver analisados a partir de descritores quantitativos e qualitativos, considerando três coleções temáticas de mandioca para os principais atributos de produtividade (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality).

Característica	Coleções				Característica	Coleções			
	CC_Completa	CC_Root_quality	CC_Yield	CC_Disease		CC_Completa	CC_Root_quality	CC_Yield	CC_Disease
AP	0,73	0,63	0,78	0,77	D_Anth	0,73	0,75	0,65	0,84
PI_Por	0,77	0,75	0,91	0,81	D_BacSev	0,76	0,68	0,85	0,85
PPA	0,77	0,61	0,86	0,82	D_MP	0,76	0,73	0,62	0,93
Ro_PRC	0,68	0,56	0,83	0,79	D_QF	0,53	0,48	0,50	0,46
Ro_PRNC	0,61	0,61	0,85	0,67	D_FerSev	0,12	0,21	0,14	0,15
PTR	0,74	0,61	0,89	0,84	D_FrogSkin	0,37	0,40	0,30	0,31
PMS	0,77	0,65	0,89	0,87	D_Pod	0,11	0,10	0,24	0,24
StC	0,37	0,71	0,56	0,34	HCN	0,90	0,47	0,85	0,87
IC	0,75	0,65	0,64	0,73	Ro_TCC	0,38	0,32	0,16	0,34
ECR	0,02	0,05	0,00	0,00	PI_Ty	0,64	0,68	0,70	0,60
CMR	0,50	0,49	0,38	0,45	PPR	0,59	0,58	0,54	0,54
DMR	0,62	0,53	0,61	0,63	CR	0,56	0,47	0,57	0,55
PI_RF	0,65	0,54	0,79	0,79	DPR	0,45	0,34	0,42	0,35
PL_Vigor12M	0,76	0,68	0,92	0,88	CCR	0,84	0,92	0,53	0,74
PI_Vigor1_5M	0,55	0,53	0,75	0,67	CER	0,63	0,41	0,51	0,63
TMS	0,66	0,77	0,69	0,64	Ro_Friab	0,57	0,78	0,57	0,56
NR	0,62	0,63	0,87	0,75	CPR	0,34	0,57	0,23	0,29
D_AcSev	0,73	0,75	0,74	0,78	Ro_Cook	0,30	0,81	0,10	0,18

AP: altura de planta; PI\_Por: porte de planta; PPA: produtividade de parte aérea; Ro\_PRC: produtividade de raízes comerciais; Ro\_PRNC: produtividade de raízes não comerciais; PTR: produtividade total de raízes; PMS: produtividade de matéria seca; StC: teor de amido; IC: índice de colheita; ECR: espessura da entrecasca (córtex) das raízes; CMR comprimento médio das raízes; DMR: diâmetro médio das raízes; PI\_RF: retenção foliar; PL\_Vigor12M: vigor das plantas avaliado aos 12 meses após o plantio; PI\_Vigor1\_5M: vigor das plantas avaliado aos 1,5 meses após o plantio; TMS: teor de matéria seca; NR: número médio de raízes por planta; D\_AcSev: severidade causada por ácaro; D\_Anth: severidade causada pela antracnose; D\_BacSev: severidade causada pela bacteriose; D\_MP: severidade mancha parda; D\_QF: queima foliar; D\_FerSev: severidade da ferrugem; D\_FrogSkin: severidade causada pelo couro de sapo;

D\_Pod: severidade causada por podridão; HCN: teor de ácido cianídrico nas raízes; Ro\_TCC: teor de carotenoides totais nas raízes; Pl\_Ty: tipo de planta; PPR: presença de pedúnculo nas raízes; CR: constrictões das raízes; DPR: destaque da casca das raízes; CCR: cor do córtex das raízes; CER: cor da casca das raízes; Ro\_Friab: avaliação da friabilidade das raízes cozidas; CPR: cor da polpa das raízes; Ro\_Cook: tempo de cozimento das raízes.

### 3.4 Análise da diversidade genética de acessos das coleções temáticas de mandioca baseada em marcadores moleculares

A maioria das coleções temáticas apresentaram parâmetros genéticos ( $H_o$ ,  $H_s$  e  $F_{is}$ ) muito semelhantes à coleção completa, em termos de médias e amplitude (Tabela 4). Além disso, as coleções temáticas conseguiram reter 90.76%, 88.10% e 88.99% dos alelos presentes na coleção completa (CC\_Disease, CC\_Root quality e CC\_Disease, respectivamente). Em relação à densidade da distribuição dos parâmetros genéticos, as principais diferenças em relação à coleção completa foram observadas na coleção temática CC\_Root quality para o parâmetro  $H_s$ , sobretudo na redução do intervalo de classes entre 0,0 e 0,2 (Figura 5). Para os demais parâmetros não foram observadas alterações muito elevadas no padrão de distribuição dos parâmetros moleculares.

**Tabela 4.** Parâmetros básicos de diversidade genética das coleções temáticas formadas com base em diferentes abordagens e informações de 20.023 marcadores SNPs.

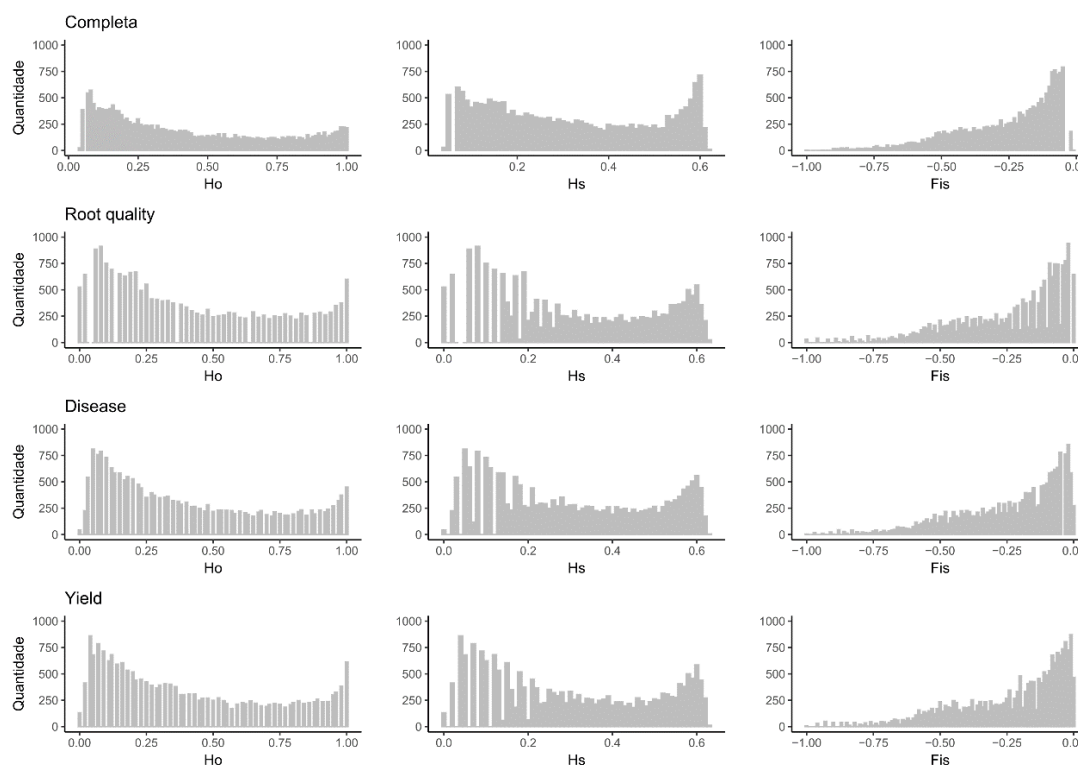
Coleções	$H_o$		$H_s$		$F_{is}$		N# total de alelo
	Média	Varição	Média	Varição	Média	Varição	
Completa	0.403	(0.04-1.00)	0.301	(0.04-0.62)	0.228	(-1.00/ 0.01)	58671
CC_Disease	0.405	(0.00-1.00)	0.302	(0.00-0.63)	-0.227	(-1.00/ 0.00)	53251
CC_Root quality	0.394	(0.00-1.00)	0.292	(0.00-0.63)	-0.226	(-1.00/ 0.00)	51688
CC_Yield	0.410	(0.00-1.00)	0.303	(0.00-0.63)	-0.234	(-1.00/ 0.00)	52211

CC\_Disease – Coleção temática para resistência a doenças e pragas; CC\_Root\_quality – Coleção temática para qualidade de raízes; CC\_Yield – Coleção temática para atributos produtivos.  $H_o$  - Heterozigosidade observada;  $H_s$  - diversidade genética dentro da população;  $F_{is}$  – Coeficiente de endogamia.

### 3.5 Validação das coleções temáticas

A PCA dos dados fenotípicos e moleculares foram utilizadas para avaliar a representatividade da diversidade de cada coleção temática e inferir a estrutura fenotípica e genética com base nas relações genéticas entre os acessos

avaliados (Figura 6). Os dois primeiros componentes principais da PCA explicaram 13,88% e 31,03% da variação dos dados moleculares e fenotípicos, respectivamente. De modo geral, a distribuição dos clones de mandioca na PCA dos dados moleculares foi maior do que os dados fenotípicos, possivelmente pela maior quantidade de dados disponíveis. Entretanto, independentemente do tipo de dado (molecular ou fenotípico) não houve um claro agrupamento dos clones selecionados para cada uma das coleções temáticas específicas, o que indica que embora existam grupos de clones de mandioca com fenótipos relativamente similares, ainda existe suficiente variabilidade genética mantida dentro das coleções temáticas para produtividade, resistência a pragas e doenças e qualidade das raízes. De certa forma essa distribuição dos clones pela PCA corrobora a diversidade molecular e fenotípica mantida nestas coleções em comparação com a coleção completa.

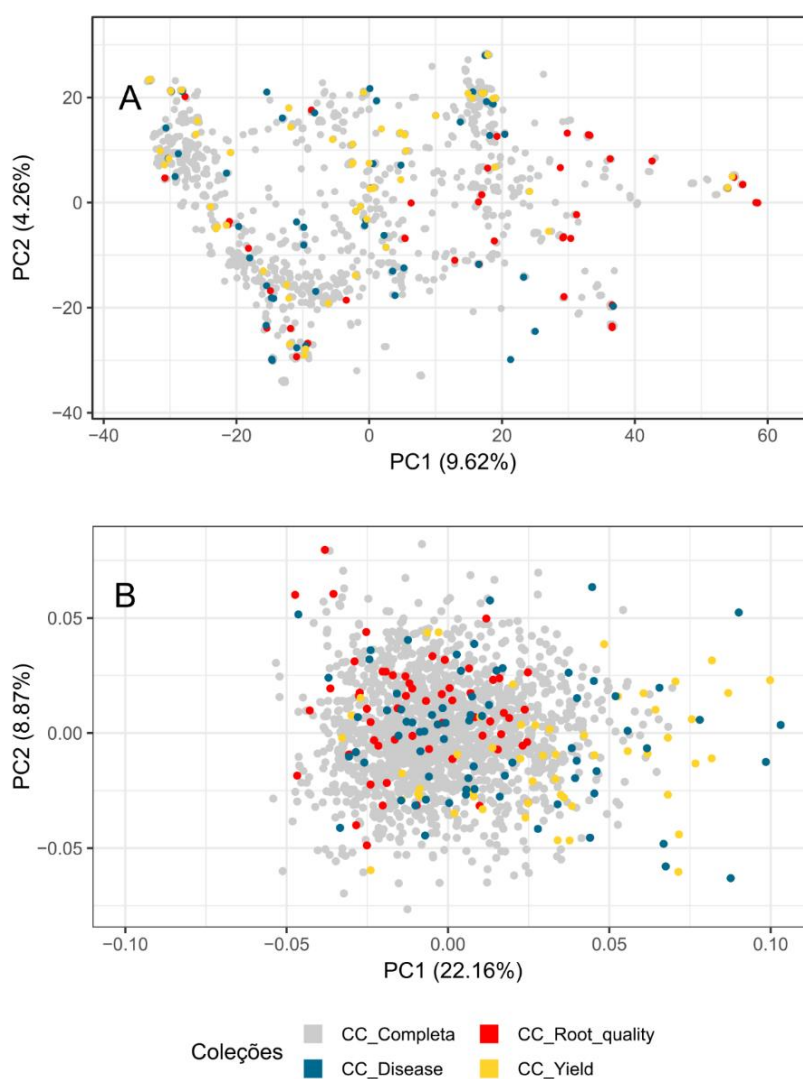


**Figura 5.** Parâmetros genéticos da coleção completa (Completa) e coleções temáticas para resistência a pragas e doenças (CC\_Disease), qualidade de raízes (CC\_Root\_quality) e atributos de produtividade (CC\_Yield). Heterozigosidade observada ( $H_o$ ); heterozigosidade esperada ( $H_s$ ) e coeficiente de endogamia ( $F_{is}$ ).

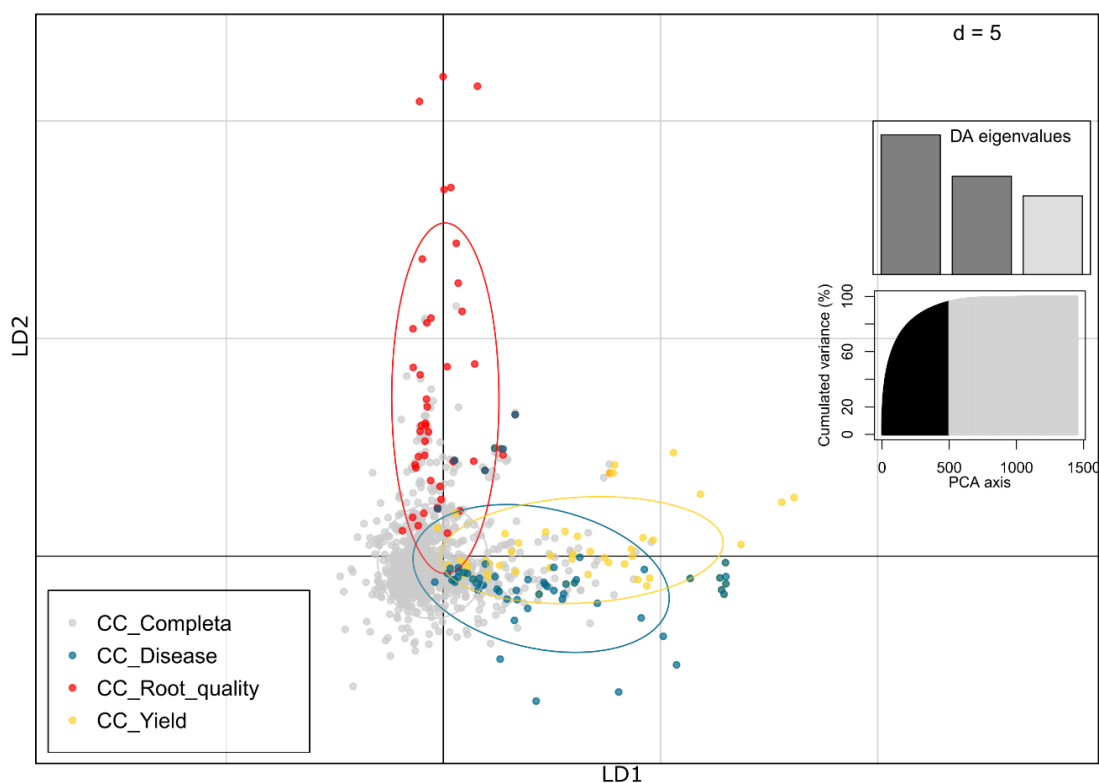
Adicionalmente utilizou-se a análise discriminante de componentes principais (ADCP) para análise conjunta dos dados moleculares e fenotípicos no intuito de verificar o padrão de distribuição dos clones da coleção temática em



comparação com a coleção completa (Figura 7). A seleção de 500 componentes principais (PC's) e três funções discriminantes (DA's) possibilitou explicar 88% da variância genética total observada. Associado a isso, a definição das coleções temáticas como grupamentos *a priori* possibilitou a discriminação dos acessos de mandioca por coleção temática. Com isso, houve uma clara separação das três coleções temáticas na ADCP, à exceção de algumas poucas sobreposições entre as coleções temáticas CC\_Disease e CC\_Yield. Por outro lado, a coleção temática CC\_Root quality foi a que permaneceu mais distanciada das demais.



**Figura 6.** Análise de componentes principais (PCA) de 1.810 acessos de mandioca com base em dados moleculares de 20.023 SNP's (A) e dados fenotípicos de 27 variáveis quantitativas (B). Os acessos em cores pertencem a coleções temáticas específicas para atributos produtivos (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality).



**Figura 7.** Análise discriminante de componentes principais (ADCP) de 20.023 SNPs e 27 descritores fenotípicos quantitativas em 1.665 acessos distribuídos diferentes coleções temáticas para atributos produtivos (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality).

## 4.0 DISCUSSÃO

### 4.1 Formação e diversidade das coleções temáticas de mandioca

Os programas de melhoramento genético de mandioca dependem da exploração eficiente da variabilidade genética armazenada nos bancos de germoplasma para seleção de parentais adequados para obtenção de clones de alto desempenho agrônomico. Entretanto, a escolha, conservação e completa caracterização dos parentais tem sido um desafio para a maioria dos programas de melhoramento de mandioca, tendo em vista os inúmeros acessos coletados e conservados nos bancos de germoplasma, além dos milhares de híbridos gerados pelos programas de melhoramento.

A estratégia de coleções nucleares, definidas como subamostras de acessos que representam a diversidade genética de uma espécie, tem sido utilizada com intuito de contornar toda essa problemática. Porém, o crescente

número e tamanho de coleções em bancos de germoplasma tem tornado necessário o estabelecimento de determinados procedimentos para limitar o tamanho das coleções. Coleções nucleares ainda podem, eventualmente, possuir um elevado número de acessos e com isso perpetuar o problema para o qual foi originalmente proposta a resolver, que seria um grupo de acessos com variabilidade genética suficiente e com um tamanho de amostra passível de ser utilizada no melhoramento genético (PESSOA FILHO, et al., 2019).

Por outro lado, coleções temáticas, compostas por acessos únicos para características específicas com tamanho administrável por programas de melhoramento tem sido uma estratégia proposta para resolver o problema de tamanho e de adequação aos interesses do melhorista (PESSOA-FILHO et al. 2010). Com base nesse propósito, o presente trabalho identificou coleções temáticas de mandioca para fins de conservação e manutenção de acessos com características úteis para obtenção de cultivares com perfil de produto adequado. Desenvolvemos assim coleções temáticas focadas em características com maior importância agrônômica para a cultura, com intuito de proceder uma extensiva caracterização agrônômica para atributos associados à aceitação comercial como mandioca de mesa e indústria, além de uso em blocos de cruzamentos para geração de progênies e obtenção dos valores de melhoramento (*breeding values*) destes clones.

Considerando que o intuito das coleções temáticas é conservar acessos que compartilham características específicas em comum, mas que ao mesmo tempo possuem características fenotípicas e alelos divergentes, é fundamental que possíveis redundâncias sejam eliminadas (MOURA et al., 2013). De fato, sabe-se que o conhecimento da relação genética entre os acessos selecionados é a primeira etapa para investigar o potencial de uso de qualquer germoplasma (EGBADZOR et al., 2014). Portanto, de acordo com o teste de IBS foram identificados 29 acessos com alto grau de parentesco genômico, que logo em seguida foram removidos das coleções temáticas ao qual eles pertenciam.

A existência de duplicatas de acessos é de ocorrência comum em bancos de germoplasmas de mandioca, e diversos estudos têm destacado o impacto negativo desses materiais nos avanços no melhoramento da espécie (MOURA et al., 2013; 2016; ALBUQUERQUE et al., 2019). Assim, ao final das análises de redundância, os acessos escolhidos de acordo a seleção truncada para as

temáticas relacionadas a atributos de produtividade (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality) foram 64, 72 e 63 acessos, respectivamente, que é um tamanho suficiente para capturar a diversidade total (fenotípica e alélica) e ainda facilitar as posteriores avaliações e validações de campo (OLIVEIRA et al., 2014).

Em soja (*Glycine max*), Guo et al. (2014) desenvolveram diversas coleções temáticas com de número de genótipos semelhantes ao nosso trabalho para atributos como tolerância ao frio, seca, salinidade, resistência ao nematoide de cisto da soja e ao vírus do mosaico da soja, além de elevado teor de proteína e gordura, com base em extensivas avaliações agrônômicas e nutricionais. Essa coleção temática foi comparada com uma segunda minicollection formada com base em marcadores microssatélites (SSR), e os resultados mostraram que a coleção temática de soja apresentou um nível de diversidade molecular similar à minicollection, mas que em nível fenotípico a coleção temática abrangeu acessos de soja com características mais desejáveis em comparação com a minicollection (GUO et al., 2014).

Embora coleções temáticas reúnam acessos com características específicas e, portanto, o foco não seja a conservação da máxima diversidade fenotípica ou genotípica, nós fizemos uma análise comparativa de diversos parâmetros genéticos com a coleção completa. Com base na análise dos marcadores moleculares as coleções temáticas CC\_Disease, CC\_Root quality e CC\_Yield, capturaram 90.76%, 88.10% e 88.99% de alelos da coleção completa, respectivamente, demonstrando uma representação alélica bastante elevada. Resultados semelhantes foram obtidos em outras espécies como arroz selvagem (*Oryza rufipogon* Griff.) na qual 90% dos alelos da coleção completa foram retidos na seleção de 130 deles para compor a coleção central de germoplasma selvagem para facilitar a conservação *ex situ* da espécie (LIU et al., 2015). De como geral a conservação da máxima variação alélica possível nas coleções temática tem relevância estratégica na manutenção da variabilidade em longo prazo porque muitas características de importância agrônômica, como resistência a doenças são frequentemente herdadas de maneira mendeliana simples (BELAJ et al., 2012).

Outro parâmetro utilizado para avaliar a qualidade das coleções temáticas foi o ISW que representa a riqueza alélica e a uniformidade dentro de uma

determinada amostra (THACHUK et al., 2009). Em nosso estudo, poucas variações entre as coleções temáticas e a coleção completa (>10%) foram observadas, demonstrando que as coleções temáticas retiveram boa parte da diversidade alélica e mantiveram o equilíbrio na distribuição das classes fenotípicas. Em outros estudos, o ISW foi um critério bastante útil para selecionar subcoleções com cobertura alélica e distância genética ótimas (BAKKALI et al., 2013).

Uma coleção de germoplasma com atributos desejáveis relacionados à qualidade de raízes, deve priorizar acessos com boas características culinárias e nutricionais. Uma das principais características limitantes para o consumo de mandioca de mesa é a presença de altos níveis de compostos cianogênicos. Assim, nossa coleção temática CC\_Root quality selecionou apenas indivíduos com  $\text{escore} < 5$  para teor de compostos cianogênicos com base na escala de classificação pelo método do picrato (BRADBURY et al., 1999), que são popularmente chamadas de “mandioca mansa” para consumo na forma cozida (BRADBURY e HOLLOWAY, 1988). Outra característica importante que foi levada em consideração na escolha dos acessos desta coleção foi o teor de carotenoides nas raízes. Além de fornecer cor amarelada nas raízes (desejada em muitas regiões no Brasil), são poderosos antioxidantes. Estes antioxidantes podem se converter em provitamina A que tem potencial de combater a fome oculta causada pela deficiência dessa vitamina de forma a contribuir para o programa de biofortificação em mandioca (CARVALHO et al., 2022a). Apesar do esforço em selecionar acessos com alto teor de carotenoides e boas características culinárias, isso não ocorreu na coleção CC\_Root quality porque a maioria dos acessos com alto teor de carotenoides também apresentaram alto teor de compostos cianogênicos. De fato, Carvalho et al. (2022b) observaram correlação positiva (0,23), embora de baixa magnitude entre teor de compostos cianogênicos e carotenoides em raízes de mandioca. Apesar disso, a coleção CC\_Root quality aumentou a frequência de clones com polpa creme ou amarela das raízes, que são características importantes para mercados que não apreciam raízes com polpa branca, independentemente do teor de carotenoides. Portanto, consideramos que as seleções feitas na coleção CC\_Root quality estão perfeitamente alinhadas com os interesses do programa de melhoramento.

Outra característica chave para a coleção temática CC\_Root quality é o

menor tempo de cozimento das raízes. A maioria dos acessos avaliados não cozinharam até 40 minutos e nossa coleção priorizou a seleção de acessos com tempo de cozimento inferior a 25 minutos, além de acessos com polpa friável ou muito friável após o cozimento. Estas duas características, são as mais importantes para a maior aceitável comercial no Brasil e em diversos países africanos, pois os consumidores usam como critério de escolha principalmente as cultivares de mesa que cozinham rapidamente e se desintegram com mais facilidade durante o cozimento (BECHOFF et al., 2018; TRAN et al., 2021).

A coleção CC\_Yield levou em consideração apenas características de interesse econômico e industrial para cultura da mandioca, cujos critérios são primordiais para que as cultivares sejam aceitas nesse mercado voltado para as agroindústrias de processamento, seja de farinha ou amido (CEBALLOS et al., 2020; ANDRADE et al., 2019). Portanto, nessa coleção temática as características com maior peso na seleção dos acessos foram a produtividade da parte aérea, raízes frescas e secas, além de características que permitem maior rendimento no processamento industrial como raízes de fácil destaque da película; raízes lisas (com poucas ou nenhuma contração) e com pedúnculo; além de plantas compactas, que permitem maior densidade de plantio e, conseqüentemente, no número de plantas por área.

Os genótipos selecionados na coleção CC\_Yield são bons candidatos para serem utilizados como parentais para melhoramento populacional, considerando a existência de alelos associados ao aumento da produtividade de raízes frescas, teor de matéria seca e amido que são os principais atributos agrônômicos para seleção em mandioca. Por outro lado, estas são características com grande influência de efeitos genéticos não-aditivos (ANDRADE et al., 2022) e por isso, é preciso levar em consideração métodos de seleção que considerem esses efeitos no melhoramento populacional e na seleção de genótipos para lançamento como cultivares. Apesar disso, a evolução de processos seletivos a exemplo da seleção genômica podem ajudar na seleção precoce de parentais e clones para avanço de geração (ANDRADE et al., 2022).

A coleção temática CC\_Disease foi direcionada para aumentar a resistência a pragas e doenças, considerando a seleção de acessos com os menores valores para severidade causada por ácaros e doenças da parte aérea

como severidade causada pela antracnose (*Colletotrichum* sp.), bacteriose (*Xanthomonas axonopodis* pv. *manihotis*), mancha parda (*Passalora henningsii*), queima foliar (*Passalora viçosae*) e ferrugem (*Uromyces manihotis*), além de podridões radiculares. Essas pragas e doenças causam redução da taxa fotossintética, danos às hastes e conseqüente redução da quantidade e qualidade do material propagativo para a próxima safra, além de penalidades na produtividade de raízes e amido (EMUOYIBOFARHE et al., 2019). Registros de perdas de mais de 75% na produtividade de raízes já foram registrados em mandioca por conta da bacteriose e antracnose (ZINSOU et al., 2004). As podridões radiculares são doenças altamente destrutivas e de difícil controle. Em diversos casos, tem-se observado perda de 100% da lavoura, causando enorme prejuízo econômico (TREMACOLDI, 2016) pelo fato de afetarem o principal produto comercial da mandioca, que são as raízes. No caso das manchas foliares e ferrugem, existem poucos registros do potencial de perda econômica, mas observações de campo indicam perdas entre 20-30% na produtividade de amido pelo fato destas doenças provocarem drástica queda de folhas, seguido de ciclos de realocação do amido das raízes para recomposição da parte aérea. Portanto, essa coleção reúne genótipos com potencial para resistência às principais doenças e patógenos da mandioca, que é a forma mais eficiente de controle e a mais adotada pelos agricultores.

## 4.2 Representação das coleções temáticas

A PCA revelou uma estrutura populacional complexa na coleção completa, sobretudo na análise dos dados moleculares. Entretanto, tanto nos dados fenotípicos quanto moleculares a PCA os acessos não apresentaram nenhum tipo de agrupamento específico em função das coleções temáticas ao qual pertenciam. Isso evidencia que os acessos de mandioca presentes em cada coleção, apesar de possuírem características comuns e desejáveis entre si, ainda são fenotípicas e genotipicamente bastante discrepantes.

De forma contrária à PCA, a ADCP usando a classificação dos acessos de mandioca em cada coleção como informação *a priori* resultou em uma melhor separação dos acessos por coleção temática. De fato, de acordo com Jombart et al. (2010), as funções discriminantes da ADCP maximizam a diversidade entre

e minimizam a diversidade dentro dos grupos. Na prática, os indivíduos e grupos são posicionados no plano usando os valores para duas variáveis quaisquer. Nesse espaço, o PCA busca a direção que apresenta a maior variância total, enquanto que a ADCP maximiza a separação entre os grupos e ainda minimiza a variação dentro do grupo. Como resultado, a PCA falha em discriminar os grupos, enquanto a ADCP exhibe adequadamente as diferenças entre os grupos (JOMBART et al., 2010). Isso explica, a separação quase que completa da coleção CC\_Root\_quality, e uma ligeira sobreposição entre as coleções CC\_Disease e CC\_Yield. Reconhecidamente, a ADCP pode ser aplicado com eficiência a uma variedade de organismos, independentemente de sua ploidia e taxas de recombinação genética, porque independe da estrutura populacional (JOMBART et al., 2010). A sensibilidade da técnica ADCP para detectar subestrutura em modelos hierárquicos em mandioca já foi demonstrada por Albuquerque et al. (2019) ao reportarem clara separação entre diferentes grupos de acessos duplicados.

### **4.3 Coleções temáticas para uso prático**

Os objetivos dos programas de melhoramento de mandioca são definidos de acordo com requerimentos específicos das diferentes regiões de recomendação e geralmente dependem de diversos atributos culturais. Entretanto, de acordo com Fukuda (2002), existem objetivos comuns em todas as regiões, como o aumento da produtividade, qualidade de raízes e resistência a pragas e doenças, por isso o presente estudo focou na construção das coleções temáticas com base nestes temas.

Com intuito de servir como material base para escolha de parentais que possam ser utilizados para o melhoramento populacional e geração de novas cultivares, mapeamento associativo, programas de seleção genômica e estudos mais refinados de transcriptoma, metiloma e epigenômica, nossos resultados tem potencial de trazer impactos mais rápidos e orientados aos melhoristas e curadores de germoplasma.

Essas coleções temáticas são um excelente ponto de partida para início do melhoramento populacional que visa aumentar a frequência dos alelos favoráveis para determinadas características agrônômicas. A modalidade da



seleção recorrente, por exemplo, é indicada para desenvolver populações superiores, principalmente naquelas características cuja herança tem controle poligênico, como é o caso da maioria das características utilizadas para a construção das coleções temáticas de mandioca.

Além do melhoramento populacional, estas coleções temáticas podem ser utilizadas para estudos de mapeamento associativo em germoplasmas ou populações biparentais visando a identificação de *quantitative trait loci* (QTL) de interesse. O mapeamento associativo foi proposto para associar alterações alélicas de uma característica de importância agrônômica a um marcador molecular (WANG et al., 2008) em coleções de indivíduos não relacionados, como uma alternativa ao mapeamento de ligação baseado em famílias (COLLARD, et al., 2005). A identificação de genes causais relacionadas a características fenotípicas específicas é um objetivo constante nos programas de melhoramento vegetal. Especificamente em mandioca, o mapeamento associativo foi utilizado para identificação de QTLs para teor de carotenoides como base para o programa de biofortificação (ESUMA et al., 2016). Wolfe et al. (2016) revelaram a arquitetura genética da resistência ao vírus do mosaico africano da mandioca causada por pelo menos quatro geminivírus do gênero *Begomovirus* (Family Geminiviridae), utilizando mapeamento associativo. Em relação a pragas, Ezenwaka et al. (2018) identificaram alelos de resistência ao ácaro verde (*Mononychellus tanajoa* (Bondar)) no germoplasma africano, utilizando um painel de diversidade de 845 genótipos melhorados de mandioca.

O desenvolvimento das coleções temáticas e triagem dos acessos com atributos de interesse para qualidade de raízes, produtividades de raízes e resistências a doenças nesse estudo, facilitará estudos de mapeamento associativo, e futuramente a possibilidade de desenvolver estudos de seleção genômica. A predição simultânea dos efeitos genéticos dos marcadores utilizados, poderá possibilitar a captura dos efeitos da maioria dos *locus*, e explicar quase que totalmente as variações genéticas das características quantitativas (RESENDE, 2008). Esta pesquisa é um primeiro passo em uma iniciativa mais ampla para caracterizar a base genética e tornar eficiente o uso do germoplasma disponível de mandioca. Assim, nosso trabalho tem potencial de desempenhar um papel vital em futuros programas de melhoramento nacionais e internacionais, especialmente relacionados as características de

produtividade e resistência a doença.

Formadas por acessos adequadas para cada temática e de tamanho reduzido, as coleções temáticas desenvolvidas no presente trabalho cumprirão o que foi proposto inicialmente nesse trabalho, que seria facilitar o gerenciamento e manuseio dos acessos mais valiosos e disponíveis em cada conjunto de germoplasma. Além disso, o foco de trabalho nestas coleções temáticas também facilitará a caracterização completa para todos os descritores disponíveis para espécie da mandioca, o que diminuirá substancialmente as redundâncias entre os acessos e aumentará as chances de identificação de alelos raros e úteis para o melhoramento da espécie.

## **5.0 CONCLUSÕES**

Este estudo desenvolveu três coleções temáticas com base em atributos de produtividade (CC-Yield), resistência a pragas e doenças (CC\_Disease) e qualidade das raízes (CC\_Root quality), formadas por cerca de 4% da coleção completa. Em cada coleção temática foi observado o deslocamento favorável das médias quando comparadas com a coleção completa, embora com valores de ISW com variação >10%, o que já era esperado devido a especificidade de cada coleção temática. A construção dessas coleções baseou-se no ranqueamento dos genótipos promissores, sendo eficiente em selecionar os mais diversos para as características de interesse em cada temática.

As três coleções temáticas são caracteristicamente não redundantes, gerenciáveis e eficientes em espelhar e reunir acessos promissores e com características de interesse para atributos de produtividade, resistências a pragas e doenças e qualidade de raízes. Essas coleções possibilitarão uma caracterização e organização mais direcionada desses acessos selecionados e, portanto, facilitando o gerenciamento e manuseio do germoplasma completo. Ademais, os acessos selecionados podem servir como material base para escolha de parentais que possam ser utilizados para melhoramento populacional e geração de novas cultivares.

## 6.0 REFERÊNCIAS

ADEYEMO O. S.; CHAVARRIAGA, P.; TOHME J.; FREGENE M.; DAVIS S. J.; SETTER T. L. Overexpression of *Arabidopsis* FLOWERING LOCUS T (FT) gene improves floral development in cassava (*Manihot esculenta* Crantz). **PLoS One**, v.12, e0181460.2017.

ALBUQUERQUE, H. Y. G. D.; OLIVEIRA, E. J.; BRITO, A. C.; ANDRADE, L. R. B.; CARMO, C. D.; MORGANTE, C. V.; FALEIRO, F. G. Identification of duplicates in cassava germplasm banks based on single-nucleotide polymorphisms (SNPs). **Scientia Agricola**, v. 76, p. 328-336, 2019.

ANDRADE, L. R. B.; SOUSA, M. B.; OLIVEIRA, E. J.; RESENDE, M. D. V.; AZEVEDO, C. F. Cassava yield traits predicted by genomic selection methods. **PLoS One**, v. 14, n. 11, p. e0224920, 2019.

ANDRADE, L. R. B.; MASSAINE e BANDEIRA, M.; WOLFE, M.; JANNINK, J-L.; RESENDE, M. D. V.; AZEVEDO, C. F.; OLIVEIRA, E. J. Increasing cassava root yield: Additive-dominant genetic models for selection of parents and clones. *Frontiers in Plant Science*, v.13, 2022, doi: 10.3389/fpls.2022.1071156

BECHOFF, A.; TOMLINS, K.; FLIEDEL, G.; BECERRA LOPEZ-LAVALLE, L. A.; WESTBY, A.; HERSHEY, C.; DUFOUR, D. Cassava traits and end-user preference: Relating traits to consumer liking, sensory perception, and genetics. **Critical Reviews in Food Science and Nutrition**, v. 58, n. 4, p. 547-567, 2018.

BELAJ, A.; DOMINGUEZ-GARCIA, M. D. C.; ATIENZA, S. G.; MARTIN URDIROZ, N.; DE LA ROSA, R.; SATOVIC, Z.; DEL RIO, C. Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. **Tree Genetics & Genomes**, v. 8, p. 365-378, 2012.

BERENDS, M. S.; LUZ, C. F.; FRIEDRICH, A. W.; SINHA, B. N.; ALBERS, C. J.; GLASNER, C. AMR: an R package for working with antimicrobial resistance data. **Journal of Statistical Software**, v. 104, p. 1-31, 2022.

BLIGHE, K.; LUN, A. PCAtools: everything principal components analysis. **R package version**, v. 2, n. 0, 2020.

BOCZKOWSKA, M.; ŁAPIŃSKI, B.; KORDULASIŃSKA, I.; DOSTATNY, D. F.; CZEMBOR, J. H. Promoting the use of common oat genetic resources through diversity analysis and core collection construction. **Plos one**, v. 11, n. 12, p. e0167855, 2016.

BOUKAR, O.; BHATTACHARJEE, R.; FATOKUN, C.; KUMAR, P.L.; GUEYE, B. Cowpea. In: Mohar, S., Upadhyaya, H.D., Bisht, I.S. (eds.). Genetic and genomic resources of grain legume improvement. **Elsevier**. p.137-156, 2013.

BRADBURY, J.H.; HOLLOWAY, W.D. Chemistry of tropical root crops: significance for nutrition and agriculture in the Pacific. **Australian Centre for International Agricultural Research**, Canberra, Australia, n.6, p.76-104, 1988.

BRADBURY, P. J.; ZHANG, Z.; KROON, D. E.; CASSTEVENS, T. M.; RAMDOSS, Y.; BUCKLER, E. S. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, n. 19, p. 2633-2635, 2007.

BRADBURY, M. G.; EGAN, S.V.; BRADBURY, J.H. Determination of all forms of cyanogens in cassava roots and cassava products using picrate paper kits. **Journal Science Food Agricultural**, v. 79, p. 593-601, 1999.

BROWNING, B. L.; BROWNING, S. R. Genotype imputation with millions of reference samples. **The American Journal of Human Genetics**, v. 98, n. 1, p. 116-126, 2016.

CARVALHO, R. R. B.; MARMOLEJO, D. F. C.; BANDEIRA E SOUSA, M.; OLIVEIRA, L. A.; OLIVEIRA, E. J. Image-based phenotyping of cassava roots for diversity studies and carotenoids prediction. **Plos one**, v. 17, n. 1, p. e0263326, 2022a.

CARVALHO, R. R. B.; BANDEIRA e SOUSA, M.; OLIVEIRA, L.A.; OLIVEIRA, E. J. Phenotypic diversity and selection in biofortified cassava germplasm for yield and quality root traits. *Euphytica*, v.218(12):173, 2022b.

CEBALLOS, H.; ROJANARIDPICHED, C.; PHUMICHAJ, C.; BECERRA, L. A.; KITTIPADAKUL, P.; IGLESIAS, C.; GRACEN, V. E. Excellence in cassava breeding: perspectives for the future. **Crop Breeding, Genetics and Genomics**, v. 2, n. 2, 2020.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and psychological measurement**, v. 20, n. 1, p. 37-46, 1960.

COLLARD, B. C.; JAHUFER, M. Z. Z.; BROUWER, J. B.; PANG, E. C. K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. **Euphytica**, v. 142, p. 169-196, 2005.

EGBADZOR, K. F.; OFORI, K.; YEBOAH, M.; ABOAGYE, L. M.; OPOKU-AGYEMAN, M. O.; DANQUAH, E. Y.; OFFEI, S. K. Diversity in 113 cowpea [*Vigna unguiculata* (L) Walp] accessions assessed with 458 SNP markers. **SpringerPlus**, v. 3, p. 1-15, 2014.

EL BAKKALI, A.; HAOUANE, H.; MOUKHLI, A.; COSTES, E.; VAN DAMME, P.; KHADARI, B. Construction of core collections suitable for association mapping to optimize use of Mediterranean olive (*Olea europaea* L.) genetic resources. **PLoS One**, v. 8, n. 5, p. e61265, 2013.

EMUOYIBOFARHE, O.; EMUOYIBOFARHE, J. O.; ADEBAYO, S.; AYANDIJI, A.; DEMEJI, O.; JAMES, O. Detection and classification of cassava diseases using machine learning. **International Journal of Computer Science and Software Engineering (IJCSSE)**, v. 8, n. 7, p. 166-176, 2019.

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

BAGUMA, Y.; KAWUKI, R. S. Genome-wide association mapping of provitamin A carotenoid content in cassava. **Euphytica**, v. 212, p. 97-110, 2016.

EZENWAKA, L.; DEL CARPIO, D. P.; JANNINK, J. L.; RABBI, I.; DANQUAH, E.; ASANTE, I.; EGESI, C. Genome-wide association study of resistance to cassava green mite pest and related traits in cassava. **Crop Science**, v. 58, n. 5, p. 1907-1918, 2018.

FRANCK, H.; CHRISTIAN, M.; NOËL, A.; BRIGITTE, P.; JOSEPH, H. D.; CORNET, D.; MATHURIN, N. C. Effects of cultivar and harvesting conditions (age, season) on the texture and taste of boiled cassava roots. **Food Chemistry**, v. 126, n. 1, p. 127-133, 2011.

FRANKEL, O.H.; BROWN, A.H.D. Plant genetic resources today: a critical appraisal. In: HOLDEN, J.H.W., WILLIAMS, J.T. (Eds.), *Crop Genetic Resources: Conservation and Evaluation*. Allen and Unwin, Winchester, p. 249–257, 1984.

FUKUDA, W. M. G.; DE OLIVEIRA, S.; IGLESIAS, C. (2002). Cassava breeding. **Crop Breeding and Applied Biotechnology**, v. 2, n. 4, 2002.

FUKUDA, W. M. G.; GUEVARA, C. L.; KAWUKI, R.; FERGUSON, M. E. Selected morphological and agronomic descriptors for the characterization of cassava. **IITA**, 2010.

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

GUO, Y.; LI, Y.; HONG, H.; QIU, L.J. Establishment of the integrated applied core collection and its comparison with mini core collection in soybean (*Glycine max*). **The Crop Journal**, v. 2, n. 1, p. 38-45, 2014.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Levantamento Sistemático da Produção Agrícola. [Acesso em: 15 março 2022.] Disponível em: <https://sidra.ibge.gov.br/tabela/>.

JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC genetics**, v. 11, n. 1, p. 1-15, 2010.

KAWANO, K.; FUKUDA, W.M.G.; CENPUKDEE, U. Genetic and environmental effects on dry matter content of cassava root. **Crop Science**, v. 26, p. 69-74, 1987.

LIU, W.; SHAHID, M. Q.; BAI, L.; LU, Z.; CHEN, Y.; JIANG, L.; LU, Y. Evaluation of genetic diversity and development of a core collection of wild rice (*Oryza rufipogon* Griff.) populations in China. **PLoS One**, v. 10, n. 12, p. e0145990, 2015.

MOURA, E. F.; FARIAS NETO, J. T. D.; SAMPAIO, J. E.; SILVA, D. T. D.; RAMALHO, G. F. Identification of duplicates of cassava accessions sampled on the North Region of Brazil using microsatellite markers. **Acta Amazonica**, v. 43, p. 461-467, 2013.

MOURA, E. F.; SOUSA, N. R.; MOURA, M. F.; DIAS, M. C.; SOUZA, E. D.; DE FARIAS NETO, J. T.; SAMPAIO, J. E. Molecular characterization of accessions of a rare genetic resource: sugary cassava (*Manihot esculenta* Crantz) from Brazilian Amazon. **Genetic Resources and Crop Evolution**, v. 63, p. 583-593, 2016.

OGBONNA, A. C.; BRAATZ DE ANDRADE, L. R.; MUELLER, L. A.; DE OLIVEIRA, E. J.; BAUCHET, G. J. Comprehensive genotyping of a Brazilian cassava (*Manihot esculenta* Crantz) germplasm bank: insights into diversification and domestication. **Theoretical and Applied Genetics**, v. 134, p. 1343-1362, 2021.

OLIVEIRA, E. J.; FERREIRA, C. F.; SANTOS, V. S.; OLIVEIRA, G. A. Development of a cassava core collection based on single nucleotide polymorphism markers. *Genet.* **Genetics and Molecular Research**, v. 13, n. 3, p. 6472-6485, 2014.

OLIVEIRA, E. J.; FUKUDA, W. M. G.; OLIVEIRA, S. A. S.; RINGENBERG, R.; SILVA, M. R.; SOUZA, A.S.; SILVA, A. L. L.; OLIVEIRA FILHO, M. S.; FARIA, M. L.; SILVA, R.J. S. BRS Novo Horizonte - a new cassava variety for industrial use. **Crop Breeding and Applied Biotechnology**, v.20, n.2., 2020.

PEREZ-FONS, L.; OVALLE, T. M.; MARUTHI, M. N.; COLVIN, J.; LOPEZ-LAVALLE, L. A. B.; FRASER, P. D. The metabotyping of an East African cassava diversity panel: A core collection for developing biotic stress tolerance in cassava. **PloS one**, v. 15, n. 11, p. e0242245, 2020.

PESSOA FILHO, M.; RANGEL, P. H. N.; FERREIRA, M. E. Extracting samples of high diversity from thematic collections of large gene banks using a genetic-distance based approach. **BMC Plant Biology**, v. 10, n. 1, p. 1-10, 2010.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing, reference index version 4.0.3. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available from, 2022.

TALMA, S. V.; ALMEIDA, S. B.; LIMA, R. M. P.; VIEIRA, H. D.; BEBERT, P. A. Tempo de cozimento e textura de raízes de mandioca. **Brazilian Journal of Food Technology**, v. 16, p. 133-138, 2013.

THACHUK, C., CROSSA, J., FRANCO, J., DREISIGACKER, S., WARBURTON, M., & DAVENPORT, G. F. Core Hunter: an algorithm for sampling genetic resources based on multiple genetic measures. **BMC bioinformatics**, v. 10, p. 1-13, 2009.

TRAN, T.; ZHANG, X.; CEBALLOS, H.; MORENO, J. L.; LUNA, J.; ESCOBAR, A.; DUFOUR, D. Correlation of cooking time with water absorption and changes in relative density during boiling of cassava roots. **International journal of food science & technology**, v. 56, n. 3, p. 1193-1205, 2021.

UPADHYAYA, H. D.; FURMAN, B. J.; DWIVEDI, S. L.; UDUPA, S. M.; GOWDA,



C. L. L.; BAUM, M.; SINGH, S. Development of a composite collection for mining germplasm possessing allelic variation for beneficial traits in chickpea. **Plant genetic resources**, v. 4, n. 1, p. 13-19, 2006.

UPADHYAYA, H. D.; VETRIVENTHAN, M.; ASIRI, A. M.; CR AZEVEDO, V.; SHARMA, H. C.; SHARMA, R.; WANG, Y. H. Multi-trait diverse germplasm sources from mini core collection for sorghum improvement. **Agriculture**, v. 9, n. 6, p. 121, 2019.

VASCONCELOS, L. M.; BRITO, A. C.; CARMO, C. D.; OLIVEIRA, P. H. G. A.; OLIVEIRA, E. J. Phenotypic diversity of starch granules in cassava germplasm. **Genetics and Molecular Research**, v. 16, n. 2, p. 1-15, 2017.

VIEIRA, E. A.; FIALHO, J. D. F.; JULIO, L. D.; CARVALHO, L. J. C. B.; CORTE, J. L. D.; RINALDI, M. M.; ANJOS, J. D. R. N. D. BRS 400 and BRS 401, sweet cassava cultivars with pink roots developed by participatory breeding. **Crop Breeding and Applied Biotechnology**, v. 19, p. 501-504, 2019.

WANG, C.; HU, S.; GARDNER, C.; LÜBBERSTEDT, T. Emerging avenues for utilization of exotic germplasm. **Trends in Plant Science**, v. 22, n. 7, p. 624-637, 2017.

WANG, J.; MCCLEAN, P. E.; LEE, K.; GOOS, R. J.; HELMS, T. Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. **Theoretical And Applied Genetics**, Berlin, v. 116, n. 6, p.777-787, 2008.

WOLFE, M. D.; RABBI, I. Y.; EGESI, C.; HAMBLIN, M.; KAWUKI, R.; KULAKOW, P.; JANNINK, J. L. Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. **The Plant Genome**, v. 9, n. 2, p. plantgenome2015.11.0118, 2016.

ZHENG, X.; LEVINE, D.; SHEN, J.; GOGARTEN, S. M.; LAURIE, C.; WEIR, B.

S. A high-performance computing toolset for relatedness and principal component analysis of SNP data. **Bioinformatics**, v. 28, n. 24, p. 3326-3328, 2012.

ZINSOU, V.; WYDRA, K.; AHOHUENDO, B.; HAU, B. Genotype x environment interactions in symptom development and yield of cassava genotypes with artificial and natural cassava bacterial blight infections. **European Journal of Plant Pathology**, v. 111, p. 217-233, 2005.

## ANEXOS

**Tabela S1.** Localização geográfica e caracterização de campos dos ensaios do Banco Ativo de germoplasma de Mandioca.

Ano	Delineamento	Localização	Cidade	Altitude	Coordenadas	Tipo de solo
2011	DBC	Embrapa	Cruz das Almas	215	12°40'36.7"S, 39°05'08.0"W	LA
	DBC	Embrapa	Cruz das Almas	210	12°40'32.6"S, 39°05'13.7"W	LA
	DBA	Embrapa	Cruz das Almas	215	12°40'36.7"S, 39°05'08.0"W	LA
2012	DBA	Embrapa	Cruz das Almas	210	12°40'32.6"S, 39°05'13.7"W	LA
	DBA	Embrapa	Cruz das Almas	200	12°40'47.4"S, 39°05'00.2"W	LA
	DBC	Coopamido	Laje	180	13°06'38.4"S, 39°16'20.4"W	LV
2013	DBC	UFRB	Cruz das Almas	210	12°39'25.9"S, 39°04'58.8"W	LA
	DBC	Embrapa	Cruz das Almas	210	12°40'32.6"S, 39°05'13.7"W	LA
	DBC	Embrapa	Cruz das Almas	217	12°40'22.8"S, 39°05'06.1"W	LA
	DBA	Embrapa	Cruz das Almas	216	12°40'22.8"S, 39°05'01.5"W	LA
	DBA	Embrapa	Cruz das Almas	216	12°40'22.8"S, 39°05'01.5"W	LA
2014	DBA	Coopamido	Laje	180	13°06'39.6"S, 39°16'17.6"W	LA
	DBA	Coopamido	Laje	175	13°06'35.6"S, 39°16'19.3"W	LA
	DBA	Embrapa	Cruz das Almas	216	12°40'22.8"S, 39°05'01.5"W	LA
	DBC	Coopamido	Laje	180	13°06'39.6"S, 39°16'17.6"W	LA
	DBA	Embrapa	Cruz das Almas	216	12°40'22.8"S, 39°05'01.5"W	LA
2015	DBA	Bahiamido	Laje	196	13°06'38.5"S, 39°16'49.0"W	LA
	DBA	Bahiamido	Valença	40	13°15'33.5"S, 39°14'12.8"W	LVA
	DBA	UFRB	Cruz das Almas	210	12°39'16.4"S, 39°04'53.4"W	LA
	DBA	Bahiamido	Laje	296	13°06'38.5"S, 39°16'49.0"W	LA
	DBA	Embrapa	Cruz das Almas	216	12°40'19.5"S, 39°05'02.5"W	LA
2016	DBA	Embrapa	Cruz das Almas	216	12°40'19.5"S, 39°05'02.5"W	LA
	DBA	Embrapa	Cruz das Almas	216	12°40'19.5"S, 39°05'02.5"W	LA
2017	DBA	Embrapa	Cruz das Almas	216	12°40'19.5"S, 39°05'02.5"W	LA
2018	DBC	UFRB	Cruz das Almas	223	12°39'51.4"S, 39°04'15.7"W	LA
	DBC	UFRB	Cruz das Almas	223	12°39'43.5"S, 39°04'12.0"W	LA
2019	DBC	UFRB	Cruz das Almas	223	12°39'43.5"S, 39°04'12.0"W	LA
	DBC	UFRB	Cruz das Almas	223	12°39'43.5"S, 39°04'12.0"W	LA
2020	DBC	UFRB	Cruz das Almas	223	12°39'49.2"S, 39°03'58.1"W	LA
	DBC	UFRB	Cruz das Almas	225	12°39'49.2"S, 39°03'58.1"W	LA
2021	DBA	UFRB	Cruz das Almas	225	12°39'49.2"S, 39°03'58.1"W	LA

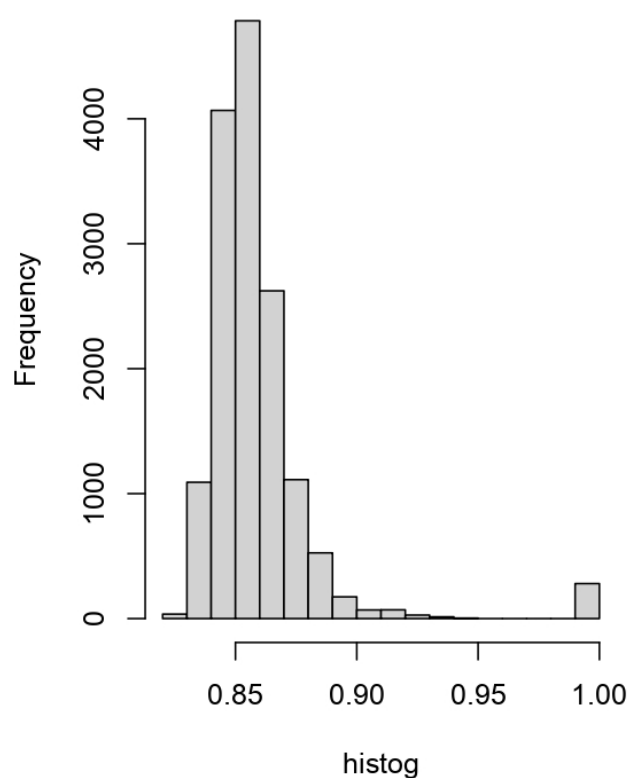
\*LA – latossolo amarelo; LV – latossolo vermelho; LVA – latossolo vermelho amarelo; DBC – Delineamento em blocos completos casualizados; DBA – Delineamento em blocos aumentados.

**Tabela S2.** Pares de indivíduos com parentesco >0.95 obtidos por IBS (*Identity by state*) utilizando 20.023 SNPs em 172 indivíduos de mandioca selecionados nas diferentes coleções temáticas (CC\_Root\_quality – Coleção temática para qualidade de raízes; CC\_Yield – Coleção temática para atributos produtivos).

1- BGM-0082/ BGM-0091	55- BGM-1474/ BGM-1507
2- BGM-0123/ BGM-0248	56- BGM-0123/ BGM-1569
3- BGM-0123 /BGM-0250	57- BGM-0248/ BGM-1569
4- BGM-0248/ BGM-0250	58- BGM-0250/ BGM-1569
5- BGM-0213/ BGM-0286	59- BGM-1227/ BGM-1569
6- BGM-0287/ BGM-0303	60- BGM-1078/ BGM -1624
7- BGM-0215/ BGM-0316	61- BGM-0093/ BGM-1626
8- BGM-0093/ BGM-0320	62- BGM-0320/ BGM-1626
9- BGM-0287/ BGM-0389	63- BGM-0666/ BGM-1638
10- BGM-0303/ BGM-0389	64- BGM-1064/ BGM-1638
11- BGM-0287/ BGM-0442	65- BGM-0189/ BGM-1643
12- BGM-0303/ BGM-0442	66- BGM-0189/ BGM-1671
13- BGM-0389/ BGM-0442	67- BGM-1643/ BGM-1671
14- BGM-0527/ BGM-0529	68- BGM-0189/ BGM-1677
15- BGM-0270/ BGM-0591	69- BGM-1643/ BGM-1677
16- BGM-0384/ BGM-0601	70- BGM-1671/ BGM-1677
17- BGM-0408/ BGM-0694	71- BGM-0123/ BGM-1679
18- BGM-0341/ BGM-0817	72- BGM-0248/ BGM-1679
19- BGM-0368/ BGM-0882	73- BGM-0250/ BGM-1679
20- BGM-0212/ BGM-1034	74- BGM-1227/ BGM-1679
21- BGM-0666/ BGM-1064	75- BGM-1569/ BGM-1679
22- BGM-0533/ BGM-1130	76- BGM-0145/ BGM-1716
23- BGM-0123/ BGM-1227	77- BGM-0145/ BGM-1721
24- BGM-0248/ BGM-1227	78- BGM-1716/ BGM-1721
25- BGM-0250/ BGM-1227	79- BGM-0368/ BGM-1817
26- BGM-0287/ BGM-1269	80- BGM-0882/ BGM-1817
27- BGM-0303/ BGM-1269	81 - BGM-0123/ BGM-1834
28- BGM-0389/ BGM-1269	82- BGM-0248/ BGM-1834
29- BGM-0442/ BGM-1269	83- BGM-0250/ BGM-1834
30- BGM-0212/ BGM-1355	84- BGM-1227/ BGM-1834
31- BGM-1034/ BGM-1355	85- BGM-1569/ BGM-1834
32- BGM-1359/ BGM-1361	86- BGM-1679/ BGM-1834
33- BGM-1359/ BGM-1371	87- BGM-0341/ BGM-1880
34- BGM-1361/ BGM-1371	88- BGM-0817/ BGM-1880
35- BGM-0443/ BGM-1395	89- BGM-1423/ BGM-1880
36- BGM-0341/ BGM-1423	90- BGM-1429/ BGM-1880
37- BGM-0817/ BGM-1423	91- BGM-1474/ BGM-1880
38- BGM-0341/ BGM-1429	92- BGM-1507/ BGM-1880
39- BGM-0817/ BGM-1429	93- BGM-0093/ BGM-2234
40- BGM-1423/ BGM-1429	94- BGM-0320/ BGM-2234

Tabela S2: continuação...

41- BGM-0307/ BGM-1465	95- BGM-1626/ BGM-2234
42- BGM-0341/ BGM-1474	96- BGM-0341/ BGM-2245
43- BGM-0817/ BGM-1474	97- BGM-0817/ BGM-2245
44- BGM-1423/ BGM-1474	98- BGM-1423/ BGM-2245
45- BGM-1429/ BGM-1474	99- BGM-1429/ BGM-2245
46- BGM-1359/ BGM-1489	100- BGM-1474/ BGM-2245
47- BGM-1361/ BGM-1489	101- BGM-1507/ BGM-2245
48- BGM-1371/ BGM-1489	102- BGM-1880/ BGM-2245
49- BGM-0579/ BGM-1498	103- BGM-1311/ BGM-2275
50- BGM-1496/ BGM-1502-M	104- BGM-2095/ BGM-2339
51- BGM-0341/ BGM-1507	105- BGM-1249/ BGM-2355
52- BGM-0817/ BGM-1507	106- BGM-2249/ BGM-2356
53- BGM-1423/ BGM-1507	107- BGM-0579/ BRS-Amansa-Burro
54- BGM-1429/ BGM-1507	108- BGM-1498/ BRS-Amansa-Burro"



**Figura S1.** Distribuição dos valores de identidade por estado alélico (IBS) entre 172 indivíduos que compõem as coleções temáticas de mandioca (CC\_Root\_quality – Coleção temática para qualidade de raízes; CC\_Yield – Coleção temática para atributos produtivos) com base na análise de 20.023 SNPs.