

**UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS, AMBIENTAIS E BIOLÓGICAS
EMBRAPA MANDIOCA E FRUTICULTURA
PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS GENÉTICOS VEGETAIS
CURSO DE MESTRADO**

**FENOTIPAGEM DIGITAL BASEADA EM VISÃO COMPUTACIONAL E
APRENDIZADO DE MÁQUINA APLICADA AO MELHORAMENTO GENÉTICO DA
MAMONEIRA**

Diego Andrade Santos

**Cruz das Almas – BAHIA
2025**

**FENOTIPAGEM DIGITAL BASEADA EM VISÃO COMPUTACIONAL E
APRENDIZADO DE MÁQUINA APLICADA AO MELHORAMENTO GENÉTICO DA
MAMONEIRA**

Diego Andrade Santos
Engenheiro Agrônomo
Universidade Federal do Recôncavo da Bahia, 2023

Dissertação apresentada ao Colegiado do Programa de Pós-Graduação em Recursos Genéticos Vegetais da Universidade Federal do Recôncavo da Bahia, como requisito parcial para obtenção do Título de Mestre em Recursos Genéticos Vegetais.

Orientador: Prof. Dr. Diego Fernando Marmolejo Cortes
Co-orientadora: Prof. Dra. Simone Alves Silva

Cruz das Almas – Bahia
2025

FICHA CATALOGRÁFICA

S237f

Santos, Diego Andrade.

Fenotipagem digital baseada em visão computacional e aprendizado de máquina aplicada ao melhoramento genético da mamoneira / Diego Andrade Santos Cruz das Almas, BA, 2025.
78f.: il.

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

Orientador: Prof. Dr. Diego Fernando Marmolejo Cortes.
Coorientadora: Prof. Dra. Simone Silva Alves.

1.Mamona – Manejo. 2.Mamona – Melhoramento genético.
3.Fenótipo – Inteligência computacional – Análise. I.Universidade Federal do Recôncavo da Bahia, Centro de Ciências Agrárias Ambientais e Biológicas. II.Título.

CDD: 633.85

Ficha elaborada pela Biblioteca Central da UFRB. Responsável pela
Elaboração Antonio Marcos Sarmiento das Chagas (Bibliotecário - CRB5 / 1615).

UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS, AMBIENTAIS E BIOLÓGICAS
EMBRAPA MANDIOCA E FRUTICULTURA
PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS GENÉTICOS VEGETAIS
CURSO DE MESTRADO

FENOTIPAGEM DIGITAL BASEADA EM VISÃO COMPUTACIONAL E
APRENDIZADO DE MÁQUINA APLICADA AO MELHORAMENTO GENÉTICO DA
MAMONEIRA

Comissão Examinadora da Defesa de Dissertação de
Diego Andrade Santos
Aprovado em: 20 de outubro de 2025

Documento assinado digitalmente
 **DIEGO FERNANDO MARMOLEJO CORTES**
Data: 20/10/2025 16:19:45-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Diego Fernando Marmolejo Cortes
Universidade Federal do Recôncavo da Bahia
(Orientador)

Documento assinado digitalmente
 **HELICIO DUARTE PEREIRA**
Data: 20/10/2025 18:56:35-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Helcio Duarte Pereira
Universidade Estadual de Campinas (Unicamp)
(Examinador Externo)

Documento assinado digitalmente
 **RENATO SANTA CATARINA**
Data: 20/10/2025 19:51:56-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Renato Santa Catarina
Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF)
(Examinador Externo)

FENOTIPAGEM DIGITAL BASEADA EM VISÃO COMPUTACIONAL E APRENDIZADO DE MÁQUINA APLICADA AO MELHORAMENTO GENÉTICO DA MAMONEIRA

RESUMO: A mamoneira (*Ricinus communis* L.) é uma oleaginosa de grande relevância industrial e estratégica, sobretudo pelo óleo extraído de suas sementes. No contexto do melhoramento genético, a caracterização fenotípica do germoplasma é essencial, porém o processo manual apresenta limitações decorrentes da subjetividade e da elevada demanda de tempo. Nesse cenário, a fenotipagem digital baseada em visão computacional surge como alternativa eficiente, ao possibilitar análises rápidas, precisas e reprodutíveis, além de apoiar a aplicação de modelos de aprendizado de máquina em programas de seleção. Este trabalho teve como objetivos: i) validar a fenotipagem digital baseada em sensores RGB como método para a caracterização morfológica de sementes de mamona, desenvolver modelos de aprendizado de máquina para a predição do peso das sementes e analisar a variabilidade genética entre acessos, visando subsidiar programas de melhoramento; e ii) caracterizar quantitativamente a coloração de folhas, galhos, pecíolos, frutos e sementes de germoplasma de mamoneira por meio de visão computacional, com aplicação em programas de melhoramento genético. Para isso, sementes e diferentes estruturas vegetais foram fotografadas em miniestúdio sob iluminação controlada. As variáveis morfológicas foram extraídas no software ImageJ®, enquanto descritores colorimétricos nos espaços de cor RGB e CIELAB foram obtidos no software R. As medições digitais de sementes foram validadas em relação a avaliações manuais e modelos de aprendizado de máquina foram aplicados para prever o peso de cem sementes (PCEM). Componentes de variância e herdabilidade foram estimados por modelos lineares mistos (REML/BLUPs) em dois anos agrícolas (2023 e 2024), e a diversidade genética foi investigada por análises multivariadas e de agrupamento. Os resultados indicaram elevada concordância entre medidas digitais e manuais pela análise de Bland-Altman e correlações superiores a 0,95. O modelo Ridge Regression apresentou maior eficiência para a predição do PCEM ($R^2 = 0,88$). A análise de componentes principais mostrou que 85,7% da variância total foi explicada pelas variáveis morfológicas das sementes, permitindo identificar agrupamentos fenotípicos consistentes entre acessos. As variáveis morfológicas Major ($H^2 = 0,88$) e AR ($H^2 = 0,87$) apresentaram alta herdabilidade, confirmando o potencial para seleção. Na caracterização colorimétrica, contrastes marcantes foram observados, como a distinção de galhos e pecíolos vermelhos em relação aos verdes e a predominância de sementes pretas (65%). As variáveis CIELAB permitiram identificar padrões específicos de coloração, com herdabilidade elevada para galhos ($H^2 > 0,90$), pecíolos ($H^2 > 0,89$) e sementes (Hue* = 1,00), evidenciando forte controle genético, enquanto as folhas apresentaram menor herdabilidade, refletindo maior influência ambiental. Em síntese, a fenotipagem digital mostrou-se ferramenta confiável e eficiente para a caracterização morfológica e colorimétrica da mamoneira, reduzindo a subjetividade das avaliações visuais, favorecendo a identificação de acessos promissores e ampliando a exploração da variabilidade genética disponível. A integração entre visão computacional e aprendizado de máquina configura uma estratégia inovadora para apoiar programas de melhoramento, ampliando descritores fenotípicos, aprimorando a seleção de acessos e fortalecendo a

conservação da diversidade em bancos de germoplasma.

Palavras-chave: Banco de germoplasma; Colorimetria; Descritores fenotípicos; Diversidade genética; Herdabilidade; Predição do peso de sementes.

COMPUTER VISION-BASED DIGITAL PHENOTYPING AND MACHINE LEARNING APPLIED TO CASTOR BEAN BREEDING

ABSTRACT: The castor bean (*Ricinus communis* L.) is an oilseed crop of great industrial and strategic relevance, mainly due to the oil extracted from its seeds. In the context of plant breeding, phenotypic characterization of germplasm is essential; however, the manual process has limitations arising from subjectivity and the high demand for time. In this scenario, digital phenotyping based on computer vision emerges as an efficient alternative, enabling rapid, precise, and reproducible analyses, as well as supporting the application of machine learning models in selection programs. This study aimed to: i) validate RGB-based digital phenotyping as a method for the morphological characterization of castor bean seeds, develop machine learning models to predict seed weight, and analyze genetic variability among accessions to support breeding programs; and ii) quantitatively characterize the color of leaves, branches, petioles, fruits, and seeds of castor bean germplasm using computer vision, with applications in genetic improvement programs. For this purpose, seeds and various plant structures were photographed in a mini-studio under controlled lighting conditions. Morphological variables were extracted using ImageJ®, while colorimetric descriptors in the RGB and CIELAB color spaces were obtained using R software. Digital seed measurements were validated against manual evaluations, and machine learning models were applied to predict hundred-seed weight (HSW). Variance components and heritability were estimated using mixed linear models (REML/BLUPs) across two growing seasons (2023 and 2024), and genetic diversity was explored through multivariate and clustering analyses. The results indicated a high agreement between digital and manual measurements, as shown by Bland–Altman analysis and correlations above 0.95. The Ridge Regression model showed the highest efficiency for HSW prediction ($R^2 = 0.88$). Principal component analysis revealed that 85.7% of the total variance was explained by seed morphological variables, allowing the identification of consistent phenotypic clusters among accessions. Morphological variables such as Major ($H^2 = 0.88$) and AR ($H^2 = 0.87$) exhibited high heritability, confirming their potential for selection. In the colorimetric characterization, marked contrasts were observed, such as the distinction between red and green branches and petioles, and the predominance of black seeds (65%). CIELAB variables enabled the identification of specific color patterns, with high heritability for branches ($H^2 > 0.90$), petioles ($H^2 > 0.89$), and seeds (Hue* = 1.00), indicating strong genetic control, while leaves showed lower heritability, reflecting greater environmental influence. In summary, digital phenotyping proved to be a reliable and efficient tool for the morphological and colorimetric characterization of castor bean, reducing subjectivity in visual assessments, facilitating the identification of promising accessions, and expanding the exploration of available genetic variability. The integration of computer vision and machine learning represents an innovative strategy to support breeding programs, broadening phenotypic descriptors, improving accession selection, and strengthening diversity conservation in germplasm banks.

Keywords: Germplasm bank; Colorimetry; Phenotypic descriptors; Genetic diversity; Heritability; Seed weight prediction.

LISTA DE TABELAS

CAPÍTULO 1

Tabela 1. Desempenho dos modelos de aprendizado de máquina na predição do peso de 100 sementes no conjunto de validação.....**35**

Tabela 2. Componentes de variância e herdabilidade para características morfológicas de sementes e rendimento de acessos do Banco ativo de Germoplasma de Mamoneira 2023 e 2024.....**39**

CAPÍTULO 2

Tabela 1. Classificação visual direta das cores do galho, pecíolo, fruto, folha e semente (cor primária) de linhagens do Banco ativo de Germoplasma de Mamoneira.....**60**

Tabela 2. Componentes de variância e herdabilidade para os parâmetros de cor (L^* , a^* , b^* , Hue* e Chroma*) em galho, pecíolo, fruto, folha e sementes de linhagens do Banco ativo de Germoplasma de Mamoneira.....**69**

LISTA DE FIGURAS

CAPÍTULO 1

Figura 1. Mini estúdio fotográfico customizado para captura de imagens.....	28
Figura 2. Processamento digital de imagens no ImageJ. A) Imagem RGB original da semente da mamona. B) Canal de cor b^* do espaço de cores Lab. C) Imagem binária. D) Imagem de contorno.....	29
Figura 3. Correlação (A–B) e concordância de Bland-Altman (C–D) entre medidas de comprimento (Comp × Major) e largura (Larg × Minor) de sementes de mamona obtidas por paquímetro e por imagens digitais.....	33
Figura 4. Correlação de Spearman entre o peso de cem sementes e características morfológicas de sementes de mamona obtidas por imagens digitais.....	34
Figura 5. Relação entre valores observados e preditos do peso de 100 sementes de mamona na fase de teste.....	36
Figura 6. A) Proporção da variância total explicada pelos componentes principais; B) contribuição das variáveis para a formação da Dimensão 1; C) contribuição das variáveis para a formação da Dimensão 2.....	37
Figura 7. Agrupamento fenotípico dos acessos do Banco Ativo de Germoplasma de Mamoneira 2023 e 2024.....	38

CAPÍTULO 2

Figura 1. Dendrogramas de agrupamento hierárquico das 65 linhagens de mamoneira obtidos a partir da análise de imagens: (A) galho, (B) pecíolo e (C) folha.....	61
Figura 2. Dendrogramas de agrupamento hierárquico das 65 linhagens de mamoneira obtidos a partir da análise de imagens: (A) Fruto, (B) semente.....	64
Figura 3. Diferença entre os grupos de linhagens de mamoneira com base em parâmetros colorimétricos cor (L^* , a^* , b^* , Hue* e Chroma*) para (A) Galho, (B) Pecíolo e (C) Folha.....	66
Figura 4. Diferença entre os grupos de linhagens de mamoneira com base em parâmetros colorimétricos cor (L^* , a^* , b^* , Hue* e Chroma*) para (A) Fruto, (B) Sementes.....	67

SUMÁRIO

1. INTRODUÇÃO GERAL	10
2. REVISÃO DE LITERATURA	12
2.1. Cultura da mamoneira	12
2.2. Melhoramento genético de mamoneira no Brasil	12
2.3. Fenotipagem de alto rendimento	13
3. REFERÊNCIAS	16
CAPÍTULO I - Fenotipagem por imagem de sementes de mamoneira para características morfológicas, predição do peso de sementes e avaliação da diversidade genética.....	21
1. INTRODUÇÃO	24
2. MATERIAL E MÉTODOS	26
2.1. Área experimental	26
2.2. Delineamento experimental.....	27
2.3. Material vegetal	27
2.4. Avaliações realizadas nas sementes	27
2.5. Aquisição das imagens.....	28
2.6. Processamento de imagens e extração de características	28
2.7. Análise estatística	29
2.7.1. Análise de Concordância pelo Método de Bland-Altman	29
2.7.2. Correlação entre Características.....	30
2.7.3. Predição de Peso de Sementes.....	30
2.7.4. Avaliação de Desempenho dos Modelos	30
2.7.5. Modelo linear misto	31
2.7.6. Análise multivariada com Componentes Principais.....	32
3. RESULTADOS	32
3.1. Validação do método por imagem	32
3.2. Correlação entre peso de cem sementes e características morfológicas.....	33
3.3. Predição do peso de cem sementes a partir de características morfológicas obtidas por imagens digitais	34
3.4. Análise de componentes principais e agrupamento fenotípico dos genótipos ..	36
3.5. Componentes de variância e herdabilidade de características morfológicas de sementes e rendimento	38
4. DISCUSSÃO	39
5. CONCLUSÕES	44
6. REFERÊNCIAS	45
CAPÍTULO II - Caracterização quantitativa da coloração de estruturas vegetais de mamoneira por visão computacional.....	51

1. INTRODUÇÃO	54
2. MATERIAL E MÉTODOS.....	55
2.1. Área experimental	55
2.2. Delineamento experimental.....	55
2.3. Material vegetal	56
2.4. Categorização visual	56
2.5. Aquisição e processamento de imagens digitais.....	56
2.6. Extração e análise do perfil quantitativo de cores	57
2.7. Obtenção de variáveis de cor derivadas	58
2.8. Análise estatística e genética.....	59
3. RESULTADOS	60
3.1. Agrupamento das cores das estruturas.....	60
3.2. Análise quantitativa das coordenadas de cor	64
3.3. Herdabilidade dos atributos de cor	68
4. DISCUSSÃO	69
5. CONCLUSÃO.....	72
6. REFERÊNCIAS.....	73

1. INTRODUÇÃO GERAL

A mamoneira (*Ricinus communis* L.) é uma cultura de destaque no cenário agrícola devido à sua importância econômica, principalmente pela produção de óleo extraído das sementes com ampla aplicação nas indústrias de lubrificantes, cosméticos e biocombustíveis (Landoni *et al.*, 2023; Ribeiro Filho *et al.*, 2023; Calerón *et al.*, 2023; Devansh; Patil; Pinjari, 2024).

No âmbito do Núcleo de Melhoramento Genético e Biotecnologia (NBIO) da Universidade Federal do Recôncavo da Bahia (UFRB), os esforços de pesquisa têm se concentrado no desenvolvimento de cultivares de mamoneira com características morfoagronômicas desejáveis, como elevado teor de óleo, alta produtividade de frutos, sementes mais pesadas, porte reduzido e adaptabilidade às condições edafoclimáticas do Nordeste brasileiro (Silva *et al.*, 2019). Para viabilizar esses avanços, o programa de melhoramento mantém um Banco Ativo de Germoplasma (BAG) de mamona composto por 89 linhagens endogâmica, cuja caracterização é realizada anualmente de forma manual por meio de descritores morfológicos que subsidiam a seleção de genótipos superiores.

Apesar de sua relevância, a caracterização morfológica de forma manual apresenta limitações operacionais, como a demanda de tempo e mão de obra, além da restrição quanto ao número de acessos avaliados simultaneamente (Bekkering; Cody; Huang, 2020). Nesse contexto, a adoção de tecnologias baseadas em sensores de imagem é promissora, possibilitando a captura precisa da variabilidade fenotípica das plantas (Jiang; Li, 2020). Essas tecnologias podem ser classificadas conforme a resolução espectral das imagens utilizadas, incluindo imagens digitais RGB (vermelho, verde e azul), multiespectrais (com três a vinte bandas espectrais) e hiperespectrais (com dezenas ou centenas de bandas), as quais fornecem dados relevantes para a caracterização fenotípica vegetal (Danilevicz *et al.*, 2021; Maciel *et al.*, 2022).

No processamento dessas imagens, ferramentas computacionais como o ImageJ® têm ampla aplicabilidade, permitindo a semi-automatização de etapas de segmentação e quantificação de características morfológicas (Schroeder *et al.*, 2020; Mazo, 2021; Elliott *et al.*, 2022).

Adicionalmente, softwares como o R ampliam as possibilidades de análise quantitativa de cores a partir de imagens digitais RGB, complementando avaliações visuais subjetivas com dados numéricos replicáveis (Weller; Westneat, 2019;

Rodríguez *et al.*, 2024). Além disso, o software R também viabiliza a aplicação de modelos preditivos baseados em aprendizado de máquina, capazes de identificar padrões em grandes volumes de dados e prever características agronômicas, como peso de sementes e produtividade (Gill *et al.*, 2022; Wang *et al.*, 2023; Sheikh *et al.*, 2024). Tais abordagens oferecem suporte à seleção de genótipos superiores, contribuindo para reduzir o tempo necessário ao desenvolvimento de novas cultivares.

Assim, este trabalho tem como objetivos: i) validar a fenotipagem digital baseada em sensores RGB como método para a caracterização morfológica de sementes de mamona, desenvolver modelos de aprendizado de máquina para a predição do peso das sementes e analisar a variabilidade genética entre acessos, visando subsidiar programas de melhoramento; e ii) caracterizar quantitativamente a coloração de folhas, galhos, pecíolos, frutos e sementes de germoplasma de mamoneira por meio de visão computacional, com aplicação em programas de melhoramento genético.

2. REVISÃO DE LITERATURA

2.1. Cultura da mamoneira

A mamoneira (*Ricinus communis* L.) é uma espécie pertencente à ordem Malpighiales, família Euphorbiaceae, gênero *Ricinus* e subtribo Ricininae (SiBBR, 2020). Trata-se de uma planta oleaginosa, de porte arbustivo, classificada como dicotiledônea, que pode apresentar ciclo de produção anual ou bianual, reproduzindo-se por meio de um sistema de polinização mista (Hussen, 2022). Suas sementes apresentam ampla variabilidade morfológica, com diferenças em formato, dimensões, coloração, peso e teor de óleo (Yeboah *et al.*, 2020).

O óleo extraído das sementes de mamona é majoritariamente composto por ácido ricinoleico, o que lhe confere propriedades físico-químicas particulares, como alta viscosidade, menor solubilidade em hexano e maior solubilidade em etanol (Chauke; Mukaya; Nkazi, 2019; Obanla *et al.*, 2021). Essas características o distinguem de outros óleos vegetais e permitem sua aplicação em diversos setores industriais, incluindo a fabricação de anticongelantes para combustíveis aeronáuticos, revestimentos automotivos, lubrificantes, resinas, tintas, cosméticos, produtos farmacêuticos e biodiesel (Vieira; Nadaleti; Sarto, 2021; Nitbani *et al.*, 2022; Landoni *et al.*, 2023).

Devido à sua importância econômica e à ampla adaptabilidade a diferentes condições edafoclimáticas, a mamoneira é cultivada em diversos países (Papazoglou *et al.*, 2020). O Brasil foi líder mundial na produção até 1980 (FAO, 2025). No entanto, a partir desse ano, o país perdeu posições no ranking e, em 2023, ocupa o segundo lugar mundial, sendo superado pela Índia, cuja produção ultrapassou a brasileira em aproximadamente 1.906.070 toneladas (FAO, 2025). Essa diferença evidencia a necessidade de avanços no setor produtivo, especialmente no campo do melhoramento genético, visando explorar o potencial agrônomo e econômico da cultura (Peixoto; Peixoto, 2025).

2.2. Melhoramento genético de mamoneira no Brasil

No Brasil, as primeiras iniciativas voltadas ao melhoramento genético da mamoneira datam de 1936, com o início dos trabalhos no Instituto Agrônomo de Campinas (IAC), em São Paulo (Krug; Mendes; Sousa, 1943). Desde então, outros programas foram desenvolvidos em instituições de ensino e pesquisa, destacando-se o Núcleo de Melhoramento Genético e Biotecnologia (NBIO), criado em 2004 na

Universidade Federal do Recôncavo da Bahia (UFRB). Esse núcleo tem se dedicado ao desenvolvimento de linhagens superiores, com características agrônômicas desejáveis, tais como porte reduzido, cápsulas semi-indeiscentes, alta produtividade, elevado teor de óleo, maior peso de sementes e resistência a pragas e doenças (Menezes *et al.*, 2021; Capinan *et al.*, 2022).

Uma das etapas essenciais desse processo envolve a conservação e a caracterização dos recursos genéticos mantidos no Banco Ativo de Germoplasma (BAG) da UFRB, que representa a base genética do programa de melhoramento (Silva *et al.*, 2019). A cada novo ciclo de plantio via semente para manutenção do BAG, realiza-se a fenotipagem dos acessos, com base em descritores qualitativos e quantitativos previamente estabelecidos para a cultura (Deepika *et al.*, 2022).

Convencionalmente, essa fenotipagem é conduzida manualmente, com base em observações visuais e medições realizadas com instrumentos, como trena e paquímetro, seguindo protocolos padronizados que permitem a comparação entre ciclos e entre genótipos (Kim *et al.*, 2022). No entanto, esse método apresenta limitações, como a subjetividade das avaliações, a influência ambiental na expressão fenotípica, a baixa precisão para caracteres complexos e a elevada demanda por tempo e mão de obra, especialmente em grandes populações de plantas (Röckel *et al.*, 2022). Para contornar essas limitações, têm sido utilizados a fenotipagem de alto rendimento, com o objetivo de aumentar a acurácia da seleção e acelerar o ganho genético (Gill *et al.*, 2022).

2.3. Fenotipagem de alto rendimento

A fenotipagem de alto rendimento é uma abordagem que utiliza sensores como RGB, multiespectrais, hiperespectrais e térmicos para a coleta rápida, precisa e eficiente de dados fenotípicos em larga escala (Abebe *et al.*, 2023). Dentre esses, o sensor RGB destaca-se por sua ampla acessibilidade, baixo custo, facilidade de operação e capacidade de captar imagens nas faixas espectrais do vermelho, verde e azul (Rezzouk *et al.*, 2020).

Os sensores RGB têm sido amplamente empregados na mensuração de características colorimétricas e morfológicas de estruturas vegetais, permitindo a construção de bancos de dados para caracterização de germoplasmas (Kim; Chung, 2021). Por exemplo, Rodríguez *et al.* (2024) utilizaram sensores RGB para extrair parâmetros de cor de vagens, sementes e flores de *Arachis* spp. e *Phaseolus* spp.,

assim como Celebioglu *et al.* (2024) que caracterizaram a variação de cor de folhas e vagens de *Phaseolus vulgaris* L., melhorando desta forma a precisão da classificação colorimétrica de acessos.

No contexto da morfologia de sementes, Kim *et al.* (2022) e Baek *et al.* (2020) utilizaram sensores RGB para obter imagens digitais de sementes de soja. Outros estudos que também utilizaram esse sensor foram realizados por Halcro *et al.* (2020), com sementes de lentilha, e Nascimento Filho *et al.* (2023), com sementes de gergelim. As imagens geradas nesses trabalhos foram processadas com o software de código aberto ImageJ®, que oferece plug-ins e ferramentas especializadas para a segmentação de regiões de interesse e a extração de características morfológicas de forma semiautomática (Schroeder *et al.*, 2020).

Além do uso de sensores e softwares, os métodos de aprendizado de máquina e estatísticos têm sido incorporados às etapas de análise fenotípica, com o objetivo de tratar grandes volumes de dados e identificar padrões complexos relacionados aos descritores avaliados (Yang *et al.*, 2024). Modelos preditivos supervisionados (treinados com dados rotulados) têm se mostrado particularmente eficazes na antecipação de características morfoagronômicas a partir de variáveis de entrada (Najafabadi; Hesami; Eskandari, 2023). Essa abordagem tem contribuído de forma significativa para a avaliação fenotípica de germoplasmas, especialmente na predição de caracteres de baixa herdabilidade ou de difícil mensuração (Silva Júnior *et al.*, 2023).

Alguns estudos abordam a aplicação dessas ferramentas em bancos de germoplasmas. Zhang *et al.* (2023), por exemplo, validaram um modelo de Multiple Linear Regression e observaram que a acurácia na predição do peso de mil sementes de trigo aumenta à medida que se eleva a correlação entre o peso e parâmetros fenotípicos das sementes, como área, perímetro, comprimento do eixo maior e comprimento do eixo menor. Em outro estudo, Duc *et al.* (2023) aplicaram sete algoritmos de aprendizado de máquina para desenvolver modelos preditivos voltados ao peso de sementes de soja. Entre os modelos avaliados nesse estudo, destacaram-se o Random Forest e Multiple Linear Regression, os quais, ao considerarem diferentes variáveis associadas às dimensões e formas das sementes, apresentaram maior precisão preditiva e menores índices de erro.

A fenotipagem de alto rendimento representa uma estratégia eficiente para otimizar os processos de seleção em programas de melhoramento genético vegetal,

viabilizando avaliações mais precisas e em grande escala, por meio do uso de sensores, softwares de análise de imagens e modelos preditivos (Omprakash *et al.*, 2024). Essa abordagem tem se destacado por contribuir para o avanço do melhoramento genético, ao aumentar a acurácia na identificação de genótipos superiores e acelerar o desenvolvimento de cultivares mais produtivas e adaptadas às diferentes condições de cultivo (Jangra *et al.*, 2021).

3. REFERÊNCIAS

- ABEBE, A.; KIM, Y.; KIM, J.; KIM, S.; BAEK, J. Image-based high-throughput phenotyping in horticultural crops. **Plants**, v. 12, n. 10, p. 2061, 2023. DOI: <https://doi.org/10.3390/plants12102061>
- BEKKERING, C. S.; HUANG, J.; TIAN, L. Image-based, organ-level plant phenotyping for wheat improvement. **Agronomy**, v. 10, n. 9, p. 1287, 2020. DOI: <https://doi.org/10.3390/agronomy10091287>
- BAEK, J.; LEE, E.; KIM, N.; KIM, S. L.; CHOI, I.; JI, H.; CHUNG, Y. S.; CHOI, M. S.; MOON, J. K.; KIM, K. H. High throughput phenotyping for various traits on soybean seeds using image analysis. **Sensors**, v. 20, n. 1, p. 248, 2020. DOI: <https://doi.org/10.3390/s20010248>
- CALERÓN, M. V.; GARCÉS, R.; FORCE, E. M.; MÉNDEZ, M. V. R.; VELASCO, L.; BARROSO, V. D.; HERRERA, C.; ALEMANY, L. J.; SALAS, J. J. High oleic castor as a new source of biodiesel 2G. **Industrial Crops and Products**, v. 204, p. 117273, 2023. DOI: <https://doi.org/10.1016/j.indcrop.2023.117273>
- CAPINAN, G. C. S.; SILVA, S. A.; SOUZA, D. R. D.; SANTOS, L. A. D.; JESUS, I. B. S. D. VARIABILITY OF F 2 PROGENIES OF CASTOR BEAN BY MEANS OF MORPHOAGRONOMIC DESCRIPTORS. **Revista Caatinga**, v. 35, p. 53-67, 2022. DOI: <https://doi.org/10.1590/1983-21252022v35n106rc>
- CELEBIOGLU, B.; MYERS, J. R.; HART, J. P.; PORCH, T.; GRIFFITHS, P. Phenotypic variability for leaf and pod color within the Snap bean association panel. **Journal of the American Society for Horticultural Science**, v. 149, n. 1, p. 15-26, 2024. DOI: 10.21273/JASHS05326-23
- CHAUKE, N. P.; MUKAYA, H. E.; NKAZI, D. B. Chemical modifications of castor oil: A review. **Science Progress**, v. 102, n. 3, p. 199-217, 2019. DOI: <https://doi.org/10.1177/0036850419859118>
- DANILEVICZ, M. F.; BAYER, P. E.; NESTOR, B. J.; BENNAMOUN, M.; EDWARDS, D. Resources for image-based high-throughput phenotyping in crops and data sharing challenges. **Plant physiology**, v. 187, n. 2, p. 699-715, 2021. DOI: <https://doi.org/10.1093/plphys/kiab301>
- DEEPIKA, C.; VENKATACHALAM S. R.; YUVARAJA, A.; RAVICHANDRAN, V.; VEERAMANI, P.; ARUTCHENTHIL, P.; INDRA, N.; KATHIRVELAN, P. Genetic analysis of diverse castor (*Ricinus communis* L.) genotypes based on seed related morphometric traits. **Electronic Journal of Plant Breeding**, v. 13, n. 4, p. 1334-1342, 2022.
- DEVANSH; PATIL, P.; PINJARI, D. V. Oil-based epoxy and their composites: A sustainable alternative to traditional epoxy. **Journal of Applied Polymer Science**, v. 141, n. 29, p. e55560, 2024. DOI: 10.1002/app.55560
- DUC, N. T.; RAMLAL, A.; RAJENDRAN, A.; RAJU, D.; LAL, S.K.; KUMAR, S.; SAHOO, R. N.; CHINNUSAMY, V. Image-based phenotyping of seed architectural

traits and prediction of seed weight using machine learning models in soybean. **Frontiers in Plant Science**, v. 14, p. 1206357, 2023. DOI: <https://doi.org/10.3389/fpls.2023.1206357>

ELLIOTT, K.; BERRY, J. C.; KIM, H.; BART, R. S. A comparison of ImageJ and machine learning based image analysis methods to measure cassava bacterial blight disease severity. **Plant Methods**, v. 18, n. 1, p. 86, 2022. DOI: <https://doi.org/10.1186/s13007-022-00906-x>

FAO. FAOSTAT: Database Results on Crops and Livestock. 2025. Disponível em: <https://www.fao.org/faostat/en/#data/QCL/visualize>. Acesso em: 28 de agosto de 2025.

GILL, T.; GILL, S. K.; SAINI, D. K.; CHOPRA, Y.; KOFF, J. P.; SANDHU, K. S. A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. **Phenomics**, v. 2, n. 3, p. 156-183, 2022. DOI: <https://doi.org/10.1007/s43657-022-00048-z>

HALCRO, K.; MCNABB, K.; LOCKINGER, A.; SOCQUET-JUGLARD, D.; BETT, K. E.; NOBLE, S. D. The BELT and phenoSEED platforms: shape and colour phenotyping of seed samples. **Plant Methods**, v. 16, p. 1-13, 2020. DOI: <https://doi.org/10.1186/s13007-020-00591-8>

HUSSEN, K. W. Review on Reproductive Biology of Caster bean (*Ricinus communis* L.). **Journal of Ecology & Natural Resources**, v. 6, n. 1, p. 000268, 2022. DOI: <https://doi.org/10.23880/jenr-16000268>

JANGRA, S.; CHAUDHARY, V.; YADAV, R. C.; YADAV, N. R. High-throughput phenotyping: A platform to accelerate crop improvement. **Phenomics**, v. 1, n. 2, p. 31-53, 2021. DOI: <https://doi.org/10.1007/s43657-020-00007-6>

JIANG, Y.; LI, C. Convolutional neural networks for image-based high-throughput plant phenotyping: a review. **Plant Phenomics**, 2020. DOI: <https://doi.org/10.34133/2020/4152816>

KIM, J. Y.; CHUNG, Y. S. A short review of RGB sensor applications for accessible high-throughput phenotyping. **Journal of Crop Science and Biotechnology**, p. 1-5, 2021. DOI: <https://doi.org/10.1007/s12892-021-00104-6>

KIM, S. H.; JO, J. W.; WANG, X.; SHIN, M. J.; HUR, O. S.; HA, B. K.; HAHN, B. S. Diversity characterization of soybean germplasm seeds using image analysis. **Agronomy**, v. 12, n. 5, p. 1004, 2022. DOI: <https://doi.org/10.3390/agronomy12051004>

KRUG, C. A.; MENDES, P. T.; SOUSA, O. Melhoramento da mamoneira (*Ricinus communis*, L.): III-primeira série de ensaios de variedades (1937/38-1938/39). **Bragantia**, v. 3, 1943. DOI: <https://doi.org/10.1590/S0006-87051943000500001>

LANDONI, M.; BERTAGNON, G.; GHIDOLI, M.; CASSANI, E.; ADANI, F.; PILU, R. Opportunities and challenges of castor bean (*Ricinus communis* L.) genetic improvement. **Agronomy**, v. 13, n. 8, p. 2076, 2023. DOI: <https://doi.org/10.3390/agronomy13082076>

MACIEL, G. M.; GALLIS, R.; BARBOSA, R. L.; CARNEIRO, V. Q.; REZENDE, W. S.; SIQUIEROLI, A. C. S. High-throughput phenotyping by RGB and multispectral imaging analysis of genotypes in sweet corn. **Horticultura Brasileira**, v. 40, p. 92-98, 2022. DOI: <https://doi.org/10.1590/s0102-0536-2022012>

MAZO, G. QuickFigures: A toolkit and ImageJ PlugIn to quickly transform microscope images into scientific figures. **PLoS One**, v. 16, n. 11, p. e0240280, 2021. DOI: <https://doi.org/10.1371/journal.pone.0240280>

MENEZES, F. D. S.; SILVA, S. A.; CAPINAN, G. C. S.; BRASILEIRO, H. S.; SANTOS, L. A. D. Identification of *Ricinus communis* hybrids for low plant height. **Revista Caatinga**, v. 34, p. 837-845, 2021. DOI: <http://dx.doi.org/10.1590/1983-21252021v34n411rc>

NAJAFABADI, M. N.; HESAMI, M.; ESKANDARI, M. Machine learning-assisted approaches in modernized plant breeding programs. **Genes**, v. 14, n. 4, p. 777, 2023. DOI: <https://doi.org/10.3390/genes14040777>

NASCIMENTO FILHO, W. B.; CIDADE, M. J. A.; PANERO, F. S.; SMIDERLE, O. J. Application of Digital Imaging Allied to Chemometrics in the Use of Non-destructive Phenotyping of Sesame Seeds. **Brazilian Journal of Analytical Chemistry**, v. 10, n. 38, p. 105-118, 2023. DOI: <http://dx.doi.org/10.30744/brjac.2179-3425.AR-88-2022>

NITBANI, F. O.; TJITDA, P. J. P.; WOGO, E. H.; DETHA, R. I. A. Preparation of Ricinoleic Acid from Castor Oil: A Review. **Journal of Oleo Science**, v. 71, n. 6, p. 781-793, 2022. DOI: <https://doi.org/10.5650/jos.ess21226>

OBANLA, O. R.; MOHAMMED, F. U.; ALEBIOSU, O. S.; OJEWUMI, M. E.; OLADIMEJI, T. E.; BABATUNDE, D. E. Study on the lubricating properties of castor (*Ricinus communis*) and hydroxylated Rubber (*Hevea brasiliensis*) seed oil. **ACS omega**, v. 6, n. 43, p. 28471-28476, 2021. DOI: <https://doi.org/10.1021/acsomega.0c05810>

OMPRAKASH; THAKUR, P.; APARNA; DOGRA, P.; BAIRWA, S. K.; MEENA, S. K.; CHANDRA K. Modern Phenotyping: A Paradigm of Present Need in Crop Improvement. **International Journal of Plant & Soil Science**, v. 36, n. 1, p. 24-37, 2024. DOI: <https://doi.org/10.9734/ijpss/2024/v36i14326>

PAPAZOGLU, E. G.; ALEXOPOULOU, E.; PAPADOPOULOS, G. K.; ECONOMOU-ANTONAKA, G. Tolerance to drought and water stress resistance mechanism of castor bean. **Agronomy**, v. 10, n. 10, p. 1580, 2020. DOI: <https://doi.org/10.3390/agronomy10101580>

PEIXOUTO, L. S.; PEIXOUTO, Y. S. Genetic Variability and Breeding Opportunities in Castor Bean (*Ricinus Communis* L.). In: ***Ricinus Communis: A Climate Resilient Commercial Crop for Sustainable Environment***. Singapore: Springer Nature Singapore, 2025. p. 35-63. DOI: https://doi.org/10.1007/978-981-96-3224-4_3

REZZOUK, F. Z.; GRACIA-ROMERO, A.; KEFAUVER, S. C.; GUTIÉRREZ, N. A.; ARANJUELO, I.; SERRET, M. D.; ARAUS, J. L. Remote sensing techniques and stable isotopes as phenotyping tools to assess wheat yield performance: Effects of growing temperature and vernalization. **Plant Science**, v. 295, p. 110281, 2020. DOI: <https://doi.org/10.1016/j.plantsci.2019.110281>

RIBEIRO FILHO, P. R. C. F.; NASCIMENTO, M. R.; SILVA, S. S. O.; LUNA, F. M. T.; CASTELLÓN, E. R.; JUNIOR, C. L. C. Synthesis and frictional characteristics of bio-based lubricants obtained from fatty acids of castor oil. **Lubricants**, v. 11, n. 2, p. 57, 2023. DOI: <https://doi.org/10.3390/lubricants11020057>

RÖCKEL, F.; SCHREIBER, T.; SCHÜLER, D.; BRAUN, U.; KRUKENBERG, I.; SCHWANDER, F.; PEIL, A.; BRANDT, C.; WILLNER, E.; GRANSOW, D.; SCHOLZ, U.; KECKE, S.; MAUL, E.; LANGE, M.; TÖPFER, R. PhenoApp: A mobile tool for plant phenotyping to record field and greenhouse observations. **F1000Research**, v. 11, p. 12, 2022. DOI: <https://doi.org/10.12688/f1000research.74239.2>

RODRÍGUEZ, D. F. C.; GUZMAN, J. J. G.; GIL, J. G. R.; WENZL, P.; URBAN, M. O. Digital descriptors sharpen classical descriptors, for improving genebank accession management: A case study on *Arachis* spp. and *Phaseolus* spp. **Plos one**, v.19, n.5, p. e0302158, 2024. DOI: <https://doi.org/10.1371/journal.pone.0302158>

SCHROEDER, A.; DOBSON, E.; RUEDEN, C.; TOMANCAK, P.; JUG, F.; ELICEIRI, K. The ImageJ ecosystem: Open-source software for image visualization, processing, and analysis. **Protein Science**, v. 30, n. 1, p. 234-249, 2020. DOI: <https://doi.org/10.1002/pro.3993>

SHEIKH, M.; IQRA, F.; AMBREEN, H.; PRAVIN, K. A.; IKRA, M.; CHUNG, Y. S. Integrating artificial intelligence and high-throughput phenotyping for crop improvement. **Journal of Integrative Agriculture**, v. 23, n. 6, p. 1787-1802, 2024. DOI: <https://doi.org/10.1016/j.jia.2023.10.019>

SiBBR - Flora do Brasil. Jardim Botânico do Rio de Janeiro. 2020. Disponível em: <http://floradobrasil.jbrj.gov.br/>. Acesso em: 28 de outubro de 2024.

SILVA, A. R.; SILVA, A. S.; SANTOS L. A.; SOUZA, D. R.; ARAUJO, G. M.; DANTAS, J. L. L.; LEITE, E. D.; DANTAS, A. C. V. L. Characterization and performance of castor bean lineages and parents at the UFRB germplasm bank. **PLoS ONE**, v. 14, n. 1, p. e0209335, 2019. DOI: <https://doi.org/10.1371/journal.pone.0209335>

SILVA JÚNIOR, A. C.; MOURA, W. M.; BHERING, L. L.; SIQUEIRA, M. J. S.; COSTA, W. G., NASCIMENTO, M.; CRUZ, C. D. Prediction and importance of predictors in approaches based on computational intelligence and machine

learning. **Agronomy Science and Biotechnology**, v. 9, p. 1-24, 2023. DOI: <https://doi.org/10.33158/ASB.r179.v9.2023>

VIEIRA, B.; NADALETI, W. C.; SARTO, E. The effect of the addition of castor oil to residual soybean oil to obtain biodiesel in Brazil: Energy matrix diversification. **Renewable Energy**, v. 165, p. 657-667, 2021. DOI: <https://doi.org/10.1016/j.renene.2020.10.056>

YANG, S.; ZHENG, L.; WU, T.; SUN, S.; ZHANG, M.; LI, M.; WANG, M. High-throughput soybean pods high-quality segmentation and seed-per-pod estimation for soybean plant breeding. **Engineering Applications of Artificial Intelligence**, v. 129, p. 107580, 2024. DOI: <https://doi.org/10.1016/j.engappai.2023.107580>

YEBOAH, A.; YING, S.; LU, J.; XIE, Y.; AMOANIMAA-DEDE, H.; BOATENG, K. G. A.; CHEN, M.; YIN, X. Castor oil (*Ricinus communis*): a review on the chemical composition and physicochemical properties. **Food Science and Technology**, v. 41, p. 399-413, 2020. DOI: <https://doi.org/10.1590/fst.19620>

WANG, W.; GUO, W.; LE, L.; YU, J.; WU, Y.; LI, D.; WANG, Y.; WANG, H.; LU, X.; QIAO, H.; GU, X.; TIAN, J.; ZHANG, C.; PU, L. Integration of high-throughput phenotyping, GWAS, and predictive models reveals the genetic architecture of plant height in maize. **Molecular Plant**, v. 16, n. 2, p. 354-373, 2023. DOI: <https://doi.org/10.1016/j.molp.2022.11.016>

WELLER, H. I.; WESTNEAT, M. W. Quantitative color profiling of digital images with earth mover's distance using the R package colordistance. **PeerJ**, v. 7, p. e6398, 2019. DOI: <https://doi.org/10.7717/peerj.6398>

ZHANG, H.; JI, J.; MA, H.; GUO, H.; LIU, N.; CUI, H. Wheat seed phenotype detection device and its application. **Agriculture**, v. 13, n. 3, p. 706, 2023. DOI: <https://doi.org/10.3390/agriculture13030706>

CAPÍTULO I

FENOTIPAGEM POR IMAGEM DE SEMENTES DE MAMONEIRA PARA CARACTERÍSTICAS MORFOLÓGICAS, PREDIÇÃO DO PESO DE SEMENTES E AVALIAÇÃO DA DIVERSIDADE GENÉTICA

FENOTIPAGEM POR IMAGEM DE SEMENTES DE MAMONEIRA PARA CARACTERÍSTICAS MORFOLÓGICAS, PREDIÇÃO DO PESO DE SEMENTES E AVALIAÇÃO DA DIVERSIDADE GENÉTICA

RESUMO: A fenotipagem digital de sementes por visão computacional é uma proposta eficiente para caracterização morfológica em programas de melhoramento genético. Este estudo teve como objetivo avaliar o potencial da fenotipagem digital para caracterização morfológica de sementes de mamoneira e sua aplicação em melhoramento genético do Núcleo de Melhoramento Genético e Biotecnologia (NBIO). As sementes de 187 acessos (ano agrícola 2023) e 124 acessos (ano agrícola 2024) do Banco Ativo de Germoplasma (BAG) de mamoneira foram fotografadas em miniestúdio com câmera RGB e processadas no software ImageJ®, utilizando macro e plugin para extração de variáveis morfológicas (Major, Minor, Área, Perímetro, AR e Round). O método digital foi validado por Bland-Altman e coeficiente de correlação. Modelos preditivos (Ridge Regression, K-Nearest Neighbors, Deep Neural Network, Random Forest e Support Vector Regression) foram treinados para prever o peso de cem sementes (PCEM). A diversidade genética foi estudada através da análise de componentes principais (PCA) e agrupamento k-means, enquanto componentes de variância e herdabilidade foram estimados usando modelos lineares mistos e BLUPs. A fenotipagem digital apresentou elevada concordância com medições manuais ($r = 0,95-0,97$) e permitiu a predição precisa do PCEM, sendo o modelo Ridge Regression o mais eficiente no conjunto de teste ($R^2 = 0,88$; RMSE = 3,83; MAE = 3,19). A PCA explicou 85,7% da variância total, revelando três agrupamentos fenotípicos distintos. Características como Major ($H^2 = 0,88$) e AR ($H^2 = 0,87$) exibiram alta herdabilidade, enquanto Round ($H^2 = 0,79$), Perímetro ($H^2 = 0,72$) e Área ($H^2 = 0,67$) apresentaram herdabilidade média. Os resultados demonstram que a fenotipagem digital é um método confiável e eficiente para caracterização de sementes de mamoneira, possibilitando a seleção de genótipos promissores e a aplicação de modelos de aprendizado de máquina para estimativas do PCEM, apoiando estratégias de melhoramento genético mais precisas e eficazes.

Palavras-chave: ImageJ, Melhoramento genético, Modelos preditivos, Visão computacional.

IMAGE-BASED PHENOTYPING OF CASTOR BEAN SEEDS FOR MORPHOLOGICAL TRAITS, SEED WEIGHT PREDICTION, AND ASSESSMENT OF GENETIC DIVERSITY

ABSTRACT: This study aimed to evaluate the potential of digital phenotyping for the morphological characterization of castor bean seeds and its application in the breeding program of the Genetic Improvement and Biotechnology Center (NBIO). Seeds from 187 accessions (2023 growing season) and 124 accessions (2024 growing season) of the Castor Bean Active Germplasm Bank (BAG) were photographed in a mini studio using an RGB camera and processed with ImageJ® software, employing macros and plugins to extract morphological variables (Major, Minor, Area, Perimeter, AR, and Round). The digital method was validated using Bland–Altman analysis and correlation coefficients. Predictive models (Ridge Regression, K-Nearest Neighbors, Deep Neural Network, Random Forest, and Support Vector Regression) were trained to predict hundred-seed weight (HSW). Genetic diversity was assessed through principal component analysis (PCA) and k-means clustering, while variance components and heritability were estimated using mixed linear models and BLUPs. Digital phenotyping showed a high agreement with manual measurements ($r = 0.95–0.97$) and allowed accurate prediction of HSW, with Ridge Regression being the most efficient model in the test set ($R^2 = 0.88$; RMSE = 3.83; MAE = 3.19). The PCA explained 85.7% of the total variance, revealing three distinct phenotypic clusters. Traits such as Major ($H^2 = 0.88$) and AR ($H^2 = 0.87$) exhibited high heritability, while Round ($H^2 = 0.79$), Perimeter ($H^2 = 0.72$), and Area ($H^2 = 0.67$) showed moderate heritability. The results demonstrate that digital phenotyping is a reliable and efficient method for characterizing castor bean seeds, enabling the selection of promising genotypes and the application of machine learning models for HSW estimation, supporting more precise and effective breeding strategies.

Keywords: ImageJ, Plant breeding, Predictive models, Computer vision.

1. INTRODUÇÃO

A mamoneira (*Ricinus communis* L.), pertencente à família Euphorbiaceae, é uma planta oleaginosa cujo óleo extraído das sementes apresenta ampla aplicação em setores industriais, incluindo a fabricação de biodiesel, cosméticos, medicamentos e plásticos biodegradáveis (Landoni *et al.*, 2023). Além do valor comercial, a espécie se destaca por sua rusticidade e capacidade de adaptação a diferentes ambientes, características que fazem do seu cultivo uma oportunidade para promover a sustentabilidade na produção agrícola e gerar benefícios econômicos para os produtores rurais (Witt *et al.*, 2021).

Considerando a relevância econômica e agrônômica da mamoneira, iniciativas de melhoramento genético tornam-se fundamentais para aumentar a produtividade da planta, bem como o peso e o teor de óleo das sementes (Delvadiya *et al.*, 2024). Nesse contexto, o Núcleo de Melhoramento Genético e Biotecnologia (NBIO) da Universidade Federal do Recôncavo da Bahia (UFRB) tem direcionado suas ações para a utilização da variabilidade genética presente no Banco Ativo de Germoplasma (BAG) de mamoneira da instituição, com o objetivo de desenvolver variedades que atendam às demandas de produtores e da indústria (Silva *et al.*, 2019).

A caracterização fenotípica de sementes dos acessos mantidos em BAGs constitui uma etapa essencial no processo de desenvolvimento de novas cultivares. O melhoramento depende da existência de variação genética nas características das sementes entre diferentes variedades, a fim de possibilitar a seleção de genótipos superiores (Siwale *et al.*, 2022). Além disso, as sementes representam a principal unidade de propagação, sendo determinantes para o estabelecimento, a produtividade e a adaptação da mamoneira (Soliman; Soliman, 2021). Entre os descritores utilizados nessa caracterização, destacam-se os atributos morfológicos das sementes, como comprimento, largura e peso, que fornecem informações sobre a diversidade entre acessos e podem estar associados a outras características agrônômicas relevantes (Zhang *et al.*, 2023).

Apesar de sua importância, a fenotipagem de sementes ainda é majoritariamente realizada por meio de métodos manuais, que demandam tempo, mão de obra especializada e elevados custos operacionais, fatores que limitam sua aplicação em larga escala (Yang *et al.*, 2021; Capinan *et al.*, 2022).

Nesse sentido, o uso de imagens tem se mostrado uma alternativa promissora para tornar a fenotipagem mais precisa, rápida e acessível (Rezzouk *et al.*, 2020). Entre os sensores disponíveis para a captura das imagens, o RGB (Red, Green, Blue) é amplamente empregado devido ao baixo custo e à capacidade de gerar imagens digitais de alta qualidade (Kim; Chung, 2021; Acosta-Navarrete *et al.*, 2023).

A utilização de sensores RGB, associada a programas de processamento de imagens, tem viabilizado medições semiautomatizadas de características morfológicas de sementes de mamoneira. Essa abordagem tem sido aplicada em estudos de diversidade genética e fenotípica, incluindo análises relacionadas à distribuição geográfica, relações filogenéticas, efeitos de fatores bioclimáticos e avaliação do vigor de sementes armazenadas (Gómez *et al.*, 2016; Elham; Soliman, 2021; Acosta-Navarrete *et al.*, 2023).

Além disso, técnicas de processamento de imagens associadas ao aprendizado de máquina têm sido empregadas principalmente para a classificação de sementes de mamoneira, tanto em materiais silvestres quanto em cultivares comerciais, utilizando algoritmos supervisionados e não supervisionados com base na frequência de índices de cores nos canais vermelho-verde-azul, matiz, saturação, intensidade e tons de cinza (Vilar *et al.*, 2015; Isaza *et al.*, 2018). Mais recentemente, o aprendizado de máquina tem avançado além da classificação, consolidando-se como ferramenta promissora também para a predição de caracteres quantitativos, como o peso de cem sementes, apresentando elevada acurácia em diferentes culturas, como arroz (Singh *et al.*, 2020), soja (Duc *et al.*, 2023; Miranda; Aono; Pinheiro, 2023) e trigo (Zhang *et al.*, 2023).

Apesar dos avanços, ainda são limitados os estudos que integram a fenotipagem de sementes por meio de processamento de imagens digitais com técnicas de aprendizado de máquina voltadas à predição do peso de cem sementes de mamoneira, bem como à investigação dos componentes genéticos das características derivadas desse processamento. Essa lacuna evidencia a necessidade de pesquisas que desenvolvam estratégias metodológicas capazes de explorar de forma mais eficiente o potencial dos recursos genéticos disponíveis.

Diante desse cenário, o objetivo do presente estudo é validar a fenotipagem digital baseada em imagens RGB para a caracterização morfológica de sementes de mamoneira, desenvolver modelos de aprendizado de máquina para a predição do

peso de sementes e analisar a variabilidade genética entre acessos, visando subsidiar programas de melhoramento da espécie.

2. MATERIAL E MÉTODOS

2.1. Área experimental

O experimento foi conduzido durante os anos agrícolas de 2023 e 2024 na área experimental do Núcleo de Melhoramento Genético e Biotecnologia (NBIO), vinculado ao Centro de Ciências Agrárias, Ambientais e Biológicas (CCAAB) da Universidade Federal do Recôncavo da Bahia (UFRB), localizada no município de Cruz das Almas, Bahia (13°12'40"S; 39°24'23"W; 220 m de altitude) (Ribeiro *et al.*, 1995).

O preparo do solo foi semelhante em ambos os anos, consistindo em aração seguida de duas gradagens. Após análise química do solo, foi realizada adubação de fundação com 15 kg ha⁻¹ de nitrogênio (N), 50 kg ha⁻¹ de fósforo (P) e 30 kg ha⁻¹ de potássio (K), aplicados diretamente nas covas (15 cm de profundidade × 20 cm de diâmetro). Adicionalmente, 20 kg/ha de N foram aplicados em cobertura.

As mudas foram produzidas em casa de vegetação, utilizando sacos de polietileno de 1,5 L, contendo substrato composto por solo argiloso e composto orgânico na proporção de 2:1 (v/v) peneirados. Devido ao baixo vigor germinativo das sementes, foi adotado o uso de mudas em substituição à semeadura no campo. Essa estratégia permitiu maior controle das condições de germinação e do desenvolvimento inicial, favorecendo a obtenção de plântulas vigorosas e aptas ao transplante. No entanto, em função da variação do vigor das sementes entre acessos, alguns não germinaram no primeiro ano agrícola, enquanto outros apresentaram germinação apenas no segundo ano.

No ano agrícola de 2023, a semeadura foi realizada em 23 de novembro e o transplante em 29 de janeiro de 2024, com espaçamento de 3 × 1 m, irrigação por gotejamento e duas aplicações de fungicida com princípio ativo Tiofanato-metílico e dosagem 60g 100L⁻¹ para controle de mofo cinzento. Em 2024, a semeadura ocorreu em 25 de abril, com transplante em 30 de maio. O espaçamento foi de 4 × 1 m, sem irrigação, e com três aplicações de inseticida com princípio ativo Deltrametrina e dosagem 156ml 100L⁻¹ de água para controlar lagarta e três aplicações de fungicida com princípio ativo Tiofanato-metílico e dosagem 60g 100L⁻¹ de água para controlar mofo-cinzento.

2.2. Delineamento experimental

Adotou-se o delineamento em blocos aumentados, com quatro blocos, para a condução do experimento. As testemunhas foram compostas pelas cultivares BRS Energia e IAC 2028, além da variedade Tomy, sendo todas repetidas nos blocos e mantidas nos dois anos agrícolas, o que possibilitou comparações entre os ciclos. Os tratamentos regulares foram representados por acessos do Banco Ativo de Germoplasma (BAG) de mamoneira, distribuídos aleatoriamente dentro dos blocos.

Em 2023, o experimento foi composto por oito linhas de plantio com 50 plantas por linha. Já em 2024, foram seis linhas com 37 plantas cada. As linhas laterais, bem como a primeira e última planta de cada linha, foram utilizadas como bordaduras.

2.3. Material vegetal

Os acessos avaliados pertencem ao BAG de mamoneira, desenvolvido a partir de hibridações controladas entre cinco genitores: BRS Nordestina, BRS Paraguaçu, EBDA MPA 34, Mirante 10 e Sipeal 28. As populações F1 foram autofecundadas para obtenção da geração F2, conduzida até as gerações F5/F6 pelo método *Single Seed Descent* (SSD), onde foi realizada a seleção dos acessos. Após mais um ciclo de autofecundação, a geração F7 foi obtida e utilizada nos experimentos.

2.4. Avaliações realizadas nas sementes

A colheita do racemo primário dos acessos, realizada nos dois anos de avaliação, foi feita com tesoura de poda após a completa dessecação dos frutos. Em seguida, foram realizadas a debulha e a quantificação do rendimento de sementes por acesso (n° de sementes planta⁻¹). No ano de 2023, também foi determinada o peso de cem sementes, a 9% de umidade, utilizando balança analítica (Shimadzu AUY220), com precisão de 0,001 g. Tanto a quantificação do rendimento quanto as análises morfológicas digitais foram conduzidas em todos os acessos avaliados em 2023 (187) e 2024 (124).

Adicionalmente, no ano agrícola de 2023, foi realizada a avaliação morfológica manual de 45 acessos, totalizando 225 sementes (cinco sementes por acesso). O comprimento (mm) e a largura (mm) dessas sementes foram

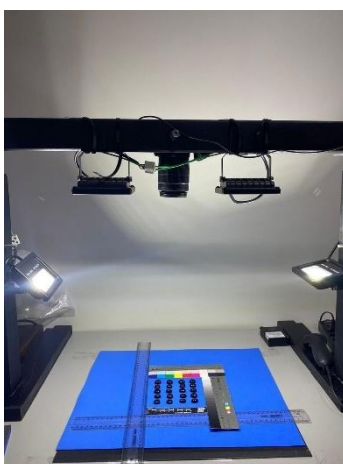
mensuradas com paquímetro digital.

2.5. Aquisição das imagens

As imagens das sementes foram capturadas com uma câmera Canon EOS REBEL T5i (lente 18–55 mm, sensor RGB), utilizando os seguintes parâmetros: velocidade do obturador 1/80 s, abertura f/5.0, ISO 200 e balanço de branco personalizado.

Para padronização da iluminação das imagens, foi utilizado um mini estúdio fotográfico customizado, composto por suporte retangular com câmera acoplada na parte superior e duas fontes de luz LED ajustáveis nas laterais (Figura 1). As sementes foram dispostas sobre fundo azul, acompanhadas por uma identificação do acesso e uma régua.

Figura 1. Mini estúdio fotográfico customizado para captura de imagens.



2.6. Processamento de imagens e extração de características

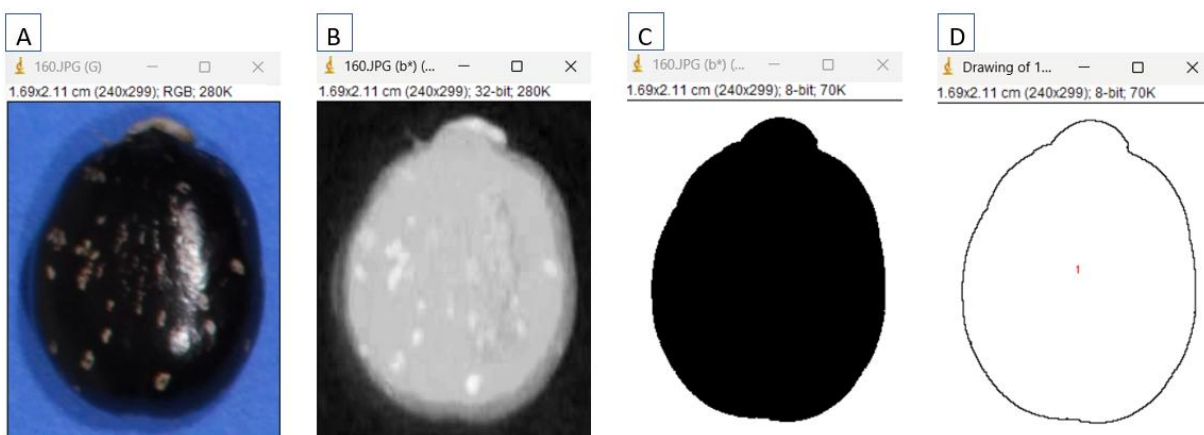
O processamento foi realizado no software ImageJ® versão 1.54g (NIH, <https://imagej.net>). As imagens RGB foram calibradas no referido software utilizando uma régua como referência de escala. Para a conversão das unidades de pixels em centímetros, foi utilizada a ferramenta “Set Scale”, adotando-se 1 cm do comprimento da régua como medida padrão. Após a calibração, as imagens foram convertidas para o espaço de cor CIELab por meio do plugin Color Space Converter. O canal b foi selecionado por apresentar melhor contraste entre as sementes e o fundo, favorecendo a definição dos contornos (Figura 2B). Para a semi-automatização do processamento, foi utilizado um macro customizado

(MacroSeedDiego), o qual realizou a conversão das imagens para escala de cinza (8 bits), aplicou a limiarização de Otsu para segmentação e, por fim, extraiu as variáveis de interesse por meio do plugin Analyze Particles (Figura 2C).

As variáveis extraídas foram:

- **Major** (diâmetro maior, mm),
- **Minor** (diâmetro menor, mm),
- **Área** (área da superfície, mm²),
- **Perímetro** (perímetro, mm),
- **AR** (razão entre diâmetro maior/diâmetro menor, $AR = \frac{\text{Eixo maior}}{\text{Eixo menor}}$)
- **Round** (índice de circularidade, $R = \frac{4 \times \text{área}}{\pi \cdot (\text{Eixo maior})^2}$).

Figura 2. Processamento digital de imagens no ImageJ. A) Imagem RGB original da semente da mamona. B) Canal de cor b* do espaço de cores Lab. C) Imagem binária. D) Imagem de contorno.



2.7. Análise estatística

2.7.1. Análise de Concordância pelo Método de Bland-Altman

A análise de concordância de Bland-Altman foi aplicada para validar o método de fenotipagem baseado em imagens. Inicialmente, foi traçada a diferença entre dois métodos de medição em relação à média dos valores obtidos pelos métodos manual e por imagens, representada em um gráfico de dispersão (Bland; Altman, 1986). No caso dos métodos de caracterização da semente, denominados *C1* e *C2*, a diferença entre as medições ($C1 - C2$) é representada no eixo y, enquanto a média das medições $(C1 + C2) / 2$ é representada no eixo x.

O viés entre os métodos, que reflete a falta de concordância, é estimado pela

diferença média ($\bar{D}$), e a variação em torno desse viés é mensurada pelo desvio padrão (σ). A variação dos resultados é calculada como $\pm 1,96 \times \sigma$, valores que definem os limites de concordância (LOA). Esses limites indicam que 95% dos pontos de dados devem estar compreendidos entre $\bar{D} \pm 1,96 \times \sigma$. A análise de concordância foi realizada por meio do pacote *pacman* (Rinker; Kurkiewicz, 2018) no software R versão 4.1.1 (R Core Team, 2023).

2.7.2. Correlação entre Características

Para examinar a associação entre as características foi realizada uma análise de correlação de Spearman implementada utilizando os pacotes *rstatix* (Kassambra, 2022) e *psych* (Revelle; Revelle, 2015) no software R versão 4.1.1 (R Core Team, 2023).

2.7.3. Predição de Peso de Sementes

Características extraídas por fenotipagem baseada em imagens de 187 acessos do BAG 2023 foram usadas para treinar e validar modelos de predição do peso de cem sementes. Os dados foram divididos em conjuntos de treinamento e teste (80-20%, respectivamente) e a performance dos modelos foi avaliada por validação cruzada. Foram avaliados seis modelos: K-Nearest Neighbors (KNN), Ridge Regression (RR), Support Vector Regression (SVR), Deep Neural Network (DNN), Random Forest (RF). A implementação dos modelos foi realizada por meio dos pacotes *caret* (Kuhn *et al.*, 2020) e *kernlab* (Karatzoglou *et al.*, 2004) no software R versão 4.1.1 (R Core Team, 2023).

2.7.4. Avaliação de Desempenho dos Modelos

O desempenho dos modelos foi avaliado com base na raiz do erro quadrático médio (RMSE) e no coeficiente de determinação (R^2), obtidos para cada partição de validação cruzada.

O RMSE mede a magnitude média dos erros estimados, sendo que valores mais próximos de 0 indicam maior precisão dos valores estimados. É calculado pela equação:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (E_i - O_i)^2}$$
 em que E_i e O_i são os valores estimados e observados, respectivamente, e n é o número de observações.

Já o R^2 varia de 0 a 1, sendo que valores mais próximos de 1 indicam maior capacidade explicativa do modelo. Representa a proporção da variação total da variável dependente explicada pela variável independente, dada pela equação:

$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$, onde $\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$ corresponde à variação explicada e $\sum_{i=1}^n (Y_i - \bar{Y})^2$ representa a variação não explicada pelo modelo.

2.7.5. Modelo linear misto

A análise genética das características de sementes obtidas através da imagem digital e do rendimento dos acessos nos dois anos agrícolas foi conduzida com base em um modelo linear misto, conforme a seguinte estrutura:

$$Y_{ijkl} = \mu + A_i + B_{(i)j} + R_k + C_l + G_m + GE_{im} + \varepsilon_{ijkl}$$

Em que Y_{ijkl} representa o valor observado do genótipo m na linha k , coluna l , e no ano i ; μ é a constante geral do experimento; A_i é o efeito fixo do ano i ; $B_{(i)j}$ é o efeito aleatório do bloco j dentro do ano i ; R_k e C_l são, respectivamente, os efeitos fixos de linha e de coluna; G_m corresponde ao efeito do genótipo, decomposto em T_m (efeito fixo do tratamento comum) e T_{jm} (efeito aleatório do tratamento regular dentro do bloco j). O efeito de interação genótipo $\times$ ano GE_{im} foi considerado aleatório e assumido como $GE_{im} \sim NID(0, \sigma_{GA}^2)$, assim como o erro experimental ($\varepsilon_{ijkl} \sim NID(0, \sigma^2)$).

A estimação dos componentes de variância foi realizada pelo método da máxima verossimilhança restrita (REML), enquanto as predições dos valores genéticos dos acessos foram obtidas por meio do estimador Best Linear Unbiased Prediction (BLUP). A significância dos efeitos fixos foi avaliada por meio do teste F, enquanto os efeitos aleatórios foram testados utilizando o Teste da Razão de Verossimilhança (LRT – *Likelihood Ratio Test*).

Os componentes de variância estimados incluíram a variância genética total (σ_g^2), a variância da interação genótipo $\times$ ambiente (σ_{ga}^2) e a variância residual (σ_e^2). A herdabilidade no sentido amplo (H^2) foi estimada com base na seguinte fórmula:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ga}^2}{a} + \frac{\sigma_e^2}{ar}}$$

em que:

- σ_g^2 é a variância genética,

- σ_{ga}^2 é a variância da interação genótipo $\times$ ambiente,
- σ_e^2 é a variância residual,
- a é o número de ambientes (anos agrícolas),
- r é o número de repetições por ambiente.

A herdabilidade foi categorizada de acordo Kim *et al.* (2025) para três classes: 0,0-40% (Baixo), 60-79% (Médio) e mais de 80% (Alto).

As análises foram realizadas no software R versão 4.4.1 (R Core Team, 2023). Para os modelos mistos e testes de significância, foram utilizados os pacotes lmerTest (Kuznetsova; Brockhoff; Christensen, 2017) e car (Fox; Weisberg, 2019).

2.7.6. Análise multivariada com Componentes Principais

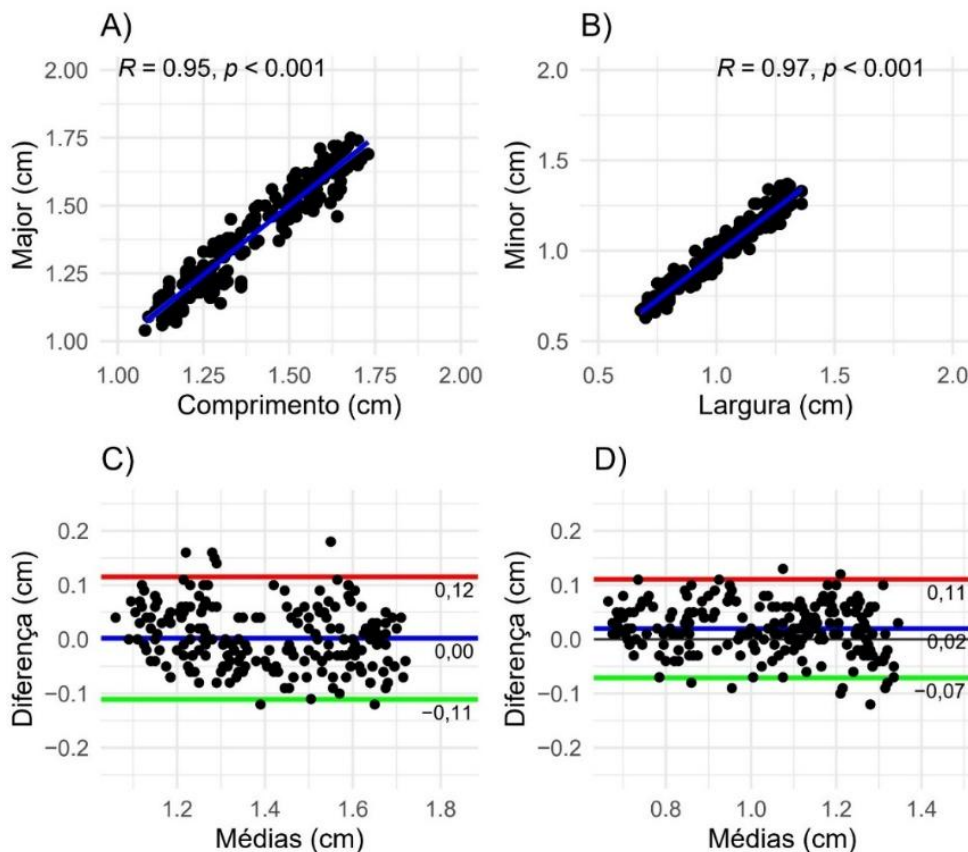
No intuito de investigar os padrões de variabilidade conjunta entre as variáveis morfológicas e a dispersão dos acessos nos dois anos avaliados, foi realizada uma Análise de Componentes Principais (PCA) utilizando os valores preditos (BLUPs). A PCA foi aplicada sobre a matriz de dados padronizada (média zero e desvio padrão igual a um), de modo a minimizar a influência de escalas distintas entre as variáveis. Para a identificação de agrupamentos entre os acessos, foi empregado o método de partição k-means com base nas coordenadas dos indivíduos nos componentes principais. A visualização gráfica foi realizada por meio de biplots gerados no software R versão 4.4.1 (R Core Team, 2023), com o auxílio dos pacotes FactoMineR (Lê; Josse; Husson, 2008) e factoextra (Kassambara; Mundt, 2017).

3. RESULTADOS

3.1. Validação do método por imagem

A comparação entre as mensurações de comprimento e diâmetro das sementes, realizadas por paquímetro e por meio de imagens digitais (em que Major e Minor correspondem, respectivamente, ao comprimento e ao diâmetro), revelou elevada concordância entre os métodos. A análise de Bland-Altman indicou viés médio de 0,0 cm para o comprimento e de 0,02 cm para o diâmetro, com as diferenças entre os pares de medidas permanecendo dentro do intervalo de 95% de concordância (Figura 3C-D). Além disso, verificou-se alta correlação entre os dados obtidos pelos dois métodos, com coeficientes de 0,95 para o comprimento e 0,97 para o diâmetro (Figura 3A-B).

Figura 3. Correlação (A–B) e concordância de Bland-Altman (C–D) entre medidas de comprimento (Comp × Major) e largura (Larg × Minor) de sementes de mamona obtidas por paquímetro e por imagens digitais.

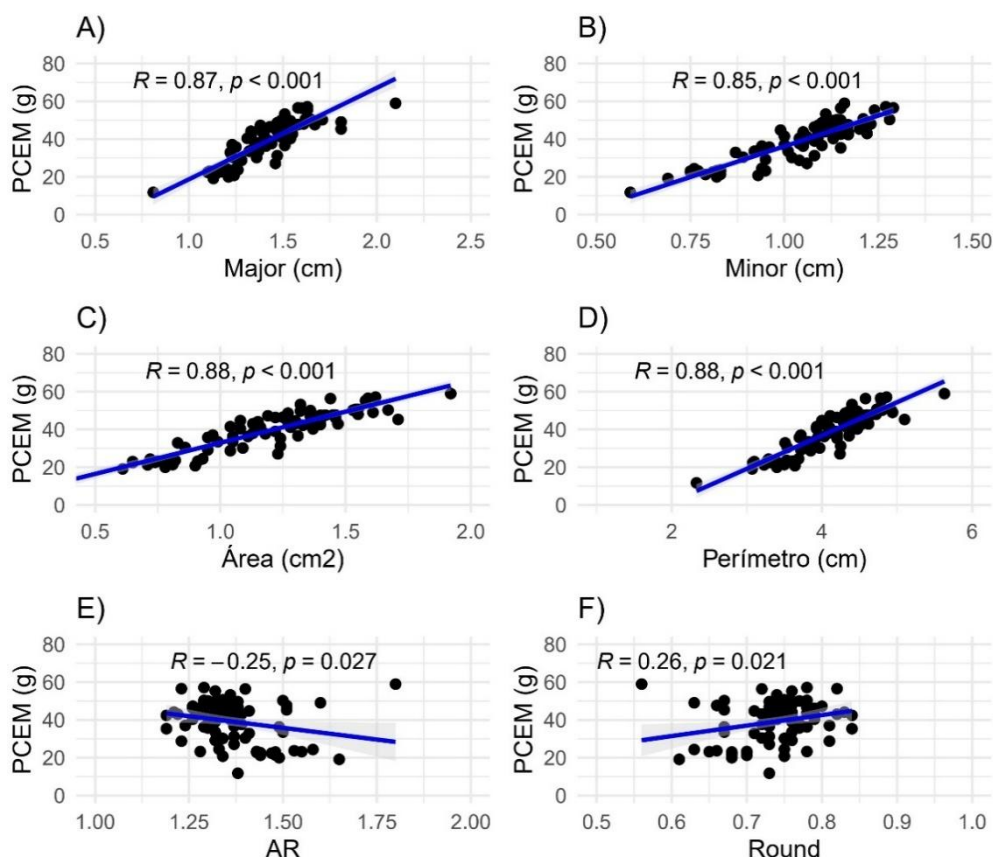


A linha azul indica o viés, e as linhas vermelha e verde os limites de concordância de 95% ($\pm 1,96 \times \sigma$).

3.2. Correlação entre peso de cem sementes e características morfológicas

O peso de cem sementes apresentou correlações elevadas e significativas ($p < 0,001$) com as características morfológicas extraídas das imagens digitais, incluindo Major, Minor, Área e Perímetro, com magnitudes de 0,87, 0,85, 0,88 e 0,88, respectivamente (Figura 4).

Figura 4. Correlação de Spearman entre o peso de cem sementes e características morfológicas de sementes de mamona obtidas por imagens digitais.



3.3. Predição do peso de cem sementes a partir de características morfológicas obtidas por imagens digitais

Modelos preditivos foram ajustados com base em diferentes algoritmos de aprendizado de máquina, utilizando variáveis morfológicas extraídas por imagens digitais (Major, Minor, Área, AR, Round e Perímetro) como preditoras do peso de 100 sementes (PCEM).

Durante a fase de treinamento, o algoritmo K-Nearest Neighbors (KNN) apresentou o melhor desempenho, com coeficiente de determinação (R^2) de 0,86 (Tabela 1). Os modelos Random Forest (RF), Ridge Regression (RR) e Deep Neural Network (DNN) também mostraram desempenho elevado, com R^2 de 0,85, 0,83 e 0,82, respectivamente. Em todos os casos, o erro quadrático médio ($RMSE < 4,84$) e absoluto médio ($MAE < 3,96$) permaneceram baixos. Por outro lado, o modelo Support Vector Regression (SVR) apresentou a menor ajuste ($R^2 = 0,72$), acompanhados dos maiores valores de erro ($RMSE = 6,13$; $MAE = 4,92$).

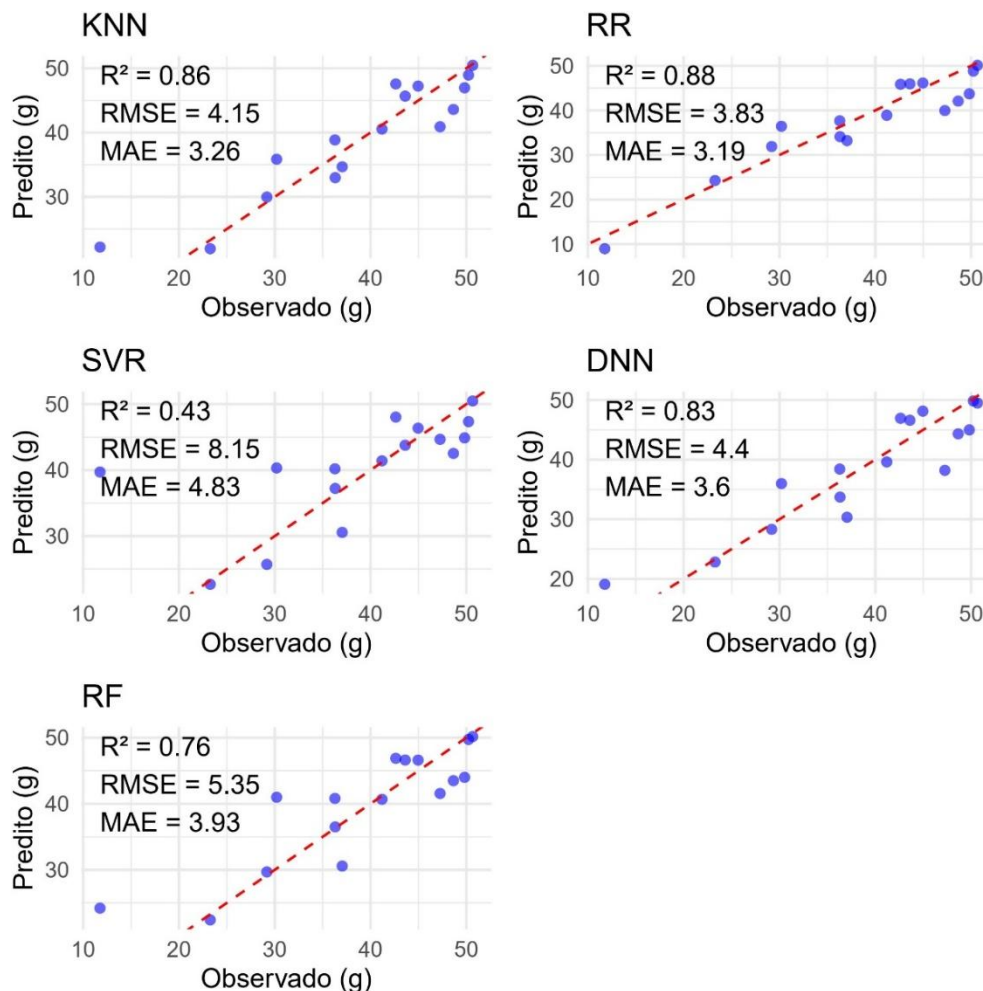
Tabela 1. Desempenho dos modelos de aprendizado de máquina na predição do peso de 100 sementes no conjunto de validação.

Modelo	R ²	RMSE	MAE
KNN	0.86	4.28	3.47
RR	0.83	4.72	3.90
SVR	0.72	6.13	4.92
DNN	0.82	4.83	3.87
RF	0.85	4.84	3.96

KNN: K-Nearest Neighbors; RR: Ridge Regression; SVR: Support Vector Regression; DNN: Deep Neural Network; RF: Random Forest. R²: coeficiente de determinação; RMSE: raiz do erro quadrático médio; MAE: erro absoluto médio.

No conjunto de teste, o modelo Ridge Regression (RR) apresentou o desempenho mais elevado, com R² de 0,88, RMSE de 3,83 e MAE de 3,19 (Figura 5). Os modelos K-Nearest Neighbor (KNN) e Deep Neural Network (DNN) também mostraram bons resultados, alcançando R² de 0,86 e 0,83, RMSE de 4,15 e 4,40, e MAE de 3,26 e 3,60, respectivamente. Em contraste, o modelo Support Vector Regression (SVR) apresentou desempenho inferior, com R² de 0,43, RMSE de 8,15 e MAE de 4,83.

Figura 5. Relação entre valores observados e preditos do peso de 100 sementes de mamona na fase de teste.

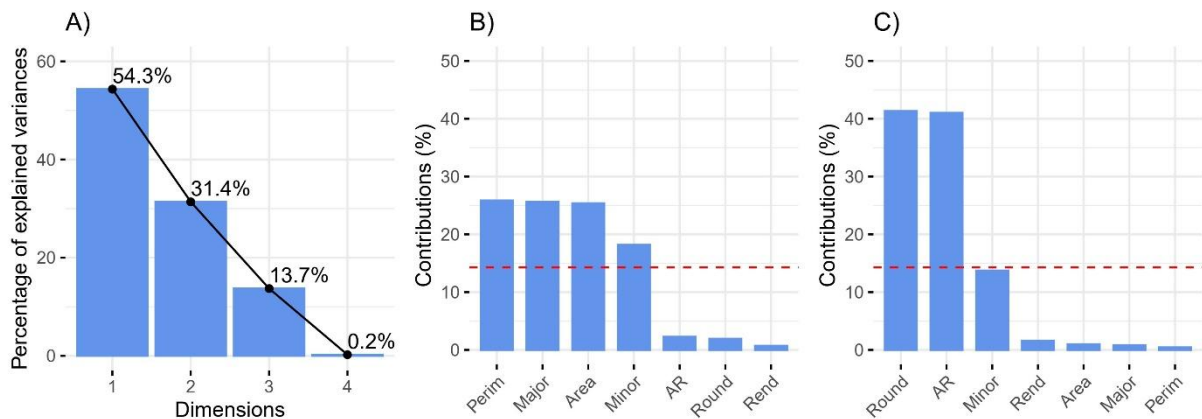


KNN: K-Nearest Neighbors; RR: Ridge Regression; SVR: Support Vector Regression; DNN: Deep Neural Network; RF: Random Forest. R^2 : coeficiente de determinação; RMSE: raiz do erro quadrático médio; MAE: erro absoluto médio.

3.4. Análise de componentes principais e agrupamento fenotípico dos genótipos

A análise de componentes principais (PCA) revelou que os dois primeiros componentes explicaram conjuntamente 85,7% da variância total, sendo 54,8% atribuídos ao primeiro componente (Dimensão 1) e 30,6% ao segundo (Dimensão 2) (Figura 6A). Em relação à contribuição das variáveis, Major, Minor, Area e Perim tiveram maior influência na Dimensão 1, enquanto AR e Round estiveram mais correlacionados à Dimensão 2. A variável Rend apresentou baixa contribuição em ambas as dimensões (Figura 6B-C).

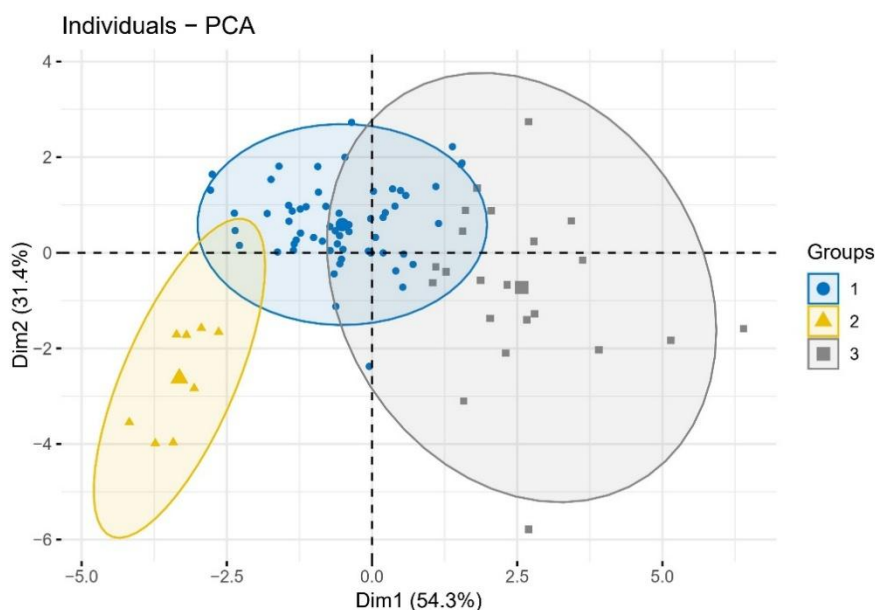
Figura 6. A) Proporção da variância total explicada pelos componentes principais; B) contribuição das variáveis para a formação da Dimensão 1; C) contribuição das variáveis para a formação da Dimensão 2.



Major (diâmetro maior), Minor (diâmetro menor), Area (área), Perímetro (perímetro), AR (razão entre diâmetro maior e diâmetro menor), Round (índice de circularidade) e Rend (rendimento).

A projeção dos acessos avaliados ao longo de dois anos agrícolas no espaço bidimensional da PCA revelou a formação de três agrupamentos fenotípicos distintos (Figura 7), definidos com base no método de agrupamento k-means aplicado aos valores de BLUPs. Os grupos diferiram principalmente em relação às características morfológicas das sementes, com o grupo 2 apresentando 8 acessos com médias de Perímetro, Major, Área e AR superiores, enquanto o grupo 1 e 3 incluíram respectivamente 64 e 23 acessos com sementes mais redondas e com maiores rendimentos.

Figura 7. Agrupamento fenotípico dos acessos do Banco Ativo de Germoplasma de Mamoneira 2023 e 2024.



Análise de Componentes Principais (PCA) utilizando os valores preditos (BLUPs) para variáveis morfológicas extraídas por imagem (Major, Minor, Área, AR, Round e Perímetro) e de rendimento, utilizando método de agrupamento k-means.

3.5. Componentes de variância e herdabilidade de características morfológicas de sementes e rendimento

A análise de deviance indicou diferenças significativas entre as características morfológicas e o rendimento em relação aos efeitos de ano, enquanto os efeitos de posição (linha, coluna e bloco) tiveram menor impacto (Tabela 2). O componente de variância associado ao ano foi significativo para Rendimento e para todas as características, exceto AR e Round.

A variância genética entre acessos foi distinta de zero para Major e AR ($p < 0,001$), Round ($p < 0,01$) e Perímetro ($p < 0,05$), evidenciando a existência de variabilidade genética explorável nessas características. Além disso, as interações entre acesso e ano ($A \times E$) foram significativas para todas as variáveis avaliadas, demonstrando a influência diferenciada do ambiente sobre o desempenho dos acessos.

As estimativas de herdabilidade no sentido amplo (H^2) variaram de 0,31 para Minor a 0,88 para Major, sendo consideradas altas para Major (0,88) e AR (0,87), intermediárias para Round (0,79), Perímetro (0,72) e Área (0,67). Esses resultados sugerem maior potencial de seleção direta para as características de alta

herdabilidade. O coeficiente de variação experimental (CV_e) apresentou valores baixos a moderados, variando de 2,22% (Round) a 10,40% (Rendimento), confirmando a boa precisão experimental.

Tabela 2. Componentes de variância e herdabilidade para características morfológicas de sementes e rendimento de acessos do Banco ativo de Germoplasma de Mamoneira 2023 e 2024.

Característica	Ensaio ^f (E)	Linha ^f	Coluna ^f	Bloco ^f	Testemunha ^f	Acesso ^a (A)	E×A ^a	Resíduo	H ²	$\bar{y}$	CV _e
	F valor	F valor	F valor	F valor	F valor	σ_g^2	σ_{ge}^2	σ^2			
Major	13,33***	2,60*	5,34**	3,54*	2,16	0,026***	0,007**	0,001	0,88	1,48	2,43
Minor	29,56***	2,38*	2,99	2,40*	7,28**	0,002	0,01***	0,001	0,31	1,09	2,56
Área	26,75***	2,68*	3,91*	2,73	4,32**	0,034	0,034***	0,003	0,67	1,30	4,52
Perímetro	30,58***	2,45*	3,80*	2,46	3,96**	0,13*	0,096***	0,011	0,72	4,35	2,38
AR	0,83	1,21	0,85	0,73	1,46	0,009***	0,002*	0,001	0,87	1,34	2,48
Round	1,19	1,35	0,93	0,65	1,54	0,002**	0,001**	0,000	0,79	0,74	2,22
Rendimento	12,37***	0,77	2,54	6,34**	0,10	697,763	1861,202***	100,654	0,43	96,44	10,40

σ_g^2 : Variância genética; σ_{ge}^2 : Variância da interação genético × ambiente; σ^2 : Variância residual; H²: herdabilidade no sentido amplo; $\bar{y}$: Média geral; CV_e : Coeficiente de variação. ^f efeito fixo; ^a efeito aleatório. Teste F para efeitos fixos e Teste da Razão de Verossimilhança (LRT) para efeitos aleatórios. ***, **, *, significativo a 0,1, 1 e 5% de probabilidade.

4. DISCUSSÃO

A fenotipagem digital de sementes por meio de visão computacional demonstrou ser uma proposta eficiente, precisa e confiável aos métodos convencionais para a mensuração de características morfológicas. Neste estudo, a concordância entre os métodos digital e manual por meio da análise de Bland-Altman associada ao alto coeficiente de correlação evidencia a viabilidade da abordagem digital. A combinação dessas análises é amplamente reconhecida na literatura como uma estratégia robusta para validar metodologias de fenotipagem de alto rendimento, conforme demonstrado por Jewan *et al.* (2022) e Plutenko *et al.* (2025), em estudos com sementes de amendoim, trigo e cevada.

Além de confiável, a fenotipagem digital aplicada neste trabalho viabilizou a avaliação de atributos morfológicos de sementes cuja mensuração manual seria inviável, como perímetro, área e índice de circularidade, além de variáveis mais comuns como diâmetro maior e diâmetro menor e a razão entre diâmetro maior e diâmetro menor (Baek *et al.*, 2020). A extração semiautomatizada dessas características foi realizada a partir de imagens obtidas com câmera digital e processadas no software de acesso livre ImageJ com auxílio de plugin e macro, o que torna o método acessível, reproduzível e dinâmico (Sunoj *et al.*, 2025).

Diferentemente de trabalhos anteriores sobre morfometria de sementes de mamoneira, como os de Isaza *et al.* (2018) e Acosta-Navarrete *et al.* (2023), que recorreram ao software proprietário Matlab, o presente estudo demonstra a viabilidade do uso de ferramentas gratuitas, ampliando as possibilidades de aplicação em contextos de pesquisa com recursos limitados.

A alta correlação de Spearman observada entre o peso de cem sementes (PCEM) e as variáveis Major, Minor, Área e Perímetro indica o potencial dessas características como preditoras diretas do PCEM (Duc *et al.*, 2023). O PCEM é utilizado em programas de melhoramento por estar associado ao rendimento e qualidade fisiológica das sementes (Marques *et al.*, 2024; Parveen *et al.*, 2025). Além disso, é utilizado como variável de interesse em estudos de mapeamento de Quantitative Trait Locus (QTLs), voltados à identificação de regiões genômicas relacionadas ao rendimento (Patel *et al.*, 2025; Chen *et al.*, 2025).

Neste estudo, foram treinados e testados modelos preditivos para PCEM a partir de atributos morfológicos extraídos digitalmente das sementes, por meio de imagens padronizadas. Os algoritmos RR, KNN e DNN apresentaram elevada acurácia preditiva, evidenciando o potencial da abordagem como alternativa à pesagem, especialmente em contextos que demandam rapidez, padronização e alta escala (El-Hendawy *et al.*, 2024). Esses modelos, embora baseados em princípios distintos, compartilham características importantes, como a aplicação em tarefas de regressão, a dependência da qualidade das variáveis preditoras e a necessidade de ajuste de hiperparâmetros, além de serem capazes de aprender padrões a partir de dados supervisionados (Kaushal *et al.*, 2024; Frías; Matínez, 2025). O bom desempenho observado pode ser atribuído à capacidade do RR em lidar com multicolinearidade, do KNN em explorar similaridades entre amostras, e do DNN em capturar relações complexas e não lineares entre as variáveis morfológicas (Jiang *et al.*, 2025; Diao *et al.*, 2025; Abekoon *et al.*, 2025).

Estudos anteriores demonstram o potencial do processamento de imagem associada a algoritmos de aprendizado de máquina, os quais têm sido empregados, principalmente para a classificação de sementes de mamoneira. Vilar *et al.* (2015) relataram elevadas taxas de classificações corretas (97,5% e 98,8%) utilizando o modelo Partial Least Squares Discriminant Analysis (PLS-DA), enquanto Isaza *et al.* (2018) obtiveram 100% de classificações corretas por meio do modelo Random Forest (RF). Esses resultados evidenciam a eficácia dessa abordagem para a

discriminação de genótipos. Entretanto, o presente estudo amplia esse escopo ao ser o primeiro, na cultura da mamoneira, a aplicar algoritmos de aprendizado de máquina não apenas para fins de classificação, mas também para a predição do PCEM. Ressalta-se que a predição do PCEM já vem sendo reportada em diferentes culturas por diversos autores, o que reforça a relevância e a inovação da aplicação nesta espécie. Duc *et al.* (2023), por exemplo, utilizaram imagens digitais processadas no SmartGrain (versão 1.2), um software de análise de imagem de código aberto comumente utilizado na extração de características de sementes, para prever o PCEM em 164 acessos de soja, empregando modelos como RF, SVR e RR, além de Simple Linear Regression (SLR), Multiple Linear Regression (MLR), LASSO e Elastic Net. O modelo RF obteve o melhor desempenho, com $R^2 = 0,98$, evidenciando alta acurácia sob as condições experimentais avaliadas. Em um estudo semelhante, Miranda, Aono e Pinheiro (2023) aplicaram AdaBoost, Perceptron Multicamada (MLP), RF e SVR para prever o PCEM em 275 genótipos de soja através de características de sementes obtidas do processamento de imagens digitais utilizando a linguagem de programação Python v.3 junto com a biblioteca OpenCV, obtendo coeficientes de determinação próximos entre os modelos (R^2 entre 0,69 e 0,72).

Diferentemente dos trabalhos com soja, Singh *et al.* (2020) desenvolveram um modelo para predição do peso e comprimento de sementes em três cultivares de arroz utilizando a contagem de pixels derivada das imagens segmentadas, e empregando uma abordagem baseada em conjunto empilhado de modelos (Stacked Ensemble Model – SEM). Na primeira camada dessa abordagem, foram utilizados algoritmos como Rede Neural Artificial (ANN), RF, SVM, Kernel Ridge Regression (KRR) e KNN, na segunda camada, o modelo RF foi empregado como meta-aprendiz. Essa abordagem resultou em alta acurácia ($R^2 = 0,95$), além de estimativas muito próximas dos valores reais de comprimento. Esses estudos reforçam que o desempenho dos modelos preditivos depende da espécie estudada, da qualidade e do volume dos dados, e dos métodos de extração das características morfológicas.

Além da predição do PCEM, as variáveis morfológicas extraídas por imagem e o rendimento do racemo primário (Rend) foram utilizadas para investigar a diversidade fenotípica presente no BAG de mamoneira nos anos agrícolas de 2023 e 2024, bem como para avaliar a significância dos componentes de variância e estimar

a herdabilidade dessas variáveis. A compreensão desses parâmetros permite aos programas de melhoramento identificar as características mais apropriadas para a seleção, contribuindo para o aumento da eficiência das estratégias para o ganho genético (Miranda; Aono; Pinheiro 2023).

A variância genética observada para os descritores Major, AR, Round e Perímetro indica que esses caracteres refletem de forma eficaz a variabilidade genética existente entre os acessos do BAG de mamoneira nas condições avaliadas, resultante da ação conjunta de efeitos gênicos aditivos e não aditivos (Zhou, 2022). Segundo Vymyslický *et al.* (2025), descritores capazes de captar adequadamente a variância genética de uma população são promissores para a caracterização de germoplasmas e para a seleção de genótipos com potencial agrônomo.

A interação significativa entre acessos e anos agrícolas ($A \times E$), observada em todas as variáveis, confirma a forte influência do ambiente sobre a expressão fenotípica. Diferenças nas condições edafoclimáticas e nas práticas de manejo entre 2023 e 2024 provavelmente contribuíram para a resposta variável dos genótipos em cada ano (Couto *et al.*, 2023). No entanto, características que, apesar da interação significativa com o ambiente, baixa variância residual e capacidade de captar a variância genética entre acessos como Major, Área, Perímetro e Round demonstram maior confiabilidade para avaliação e seleção (Figàs *et al.*, 2018).

A elevada estimativa de herdabilidade obtida para o caráter Major (0,88), nos dois anos agrícolas avaliados, corrobora os resultados obtidos em soja por Duc *et al.* (2023), que observaram herdabilidade de 0,92 em um único ambiente utilizando acessos com ampla diversidade, e por Kim *et al.* (2025), que reportaram herdabilidade de 0,94 em três ambientes distintos em uma população de acessos com ampla variabilidade genética. Esses resultados indicam que o caráter apresenta reduzida influência ambiental sobre sua expressão fenotípica e, consequentemente, elevada confiabilidade para seleção, inclusive nas fases iniciais do melhoramento, quando ainda se avalia um número elevado de indivíduos sob menor controle experimental (Roka *et al.*, 2024).

A análise de componentes principais (PCA) demonstrou ser uma ferramenta eficaz para capturar a diversidade fenotípica entre os acessos do BAG de mamoneira. Os dois primeiros componentes explicaram conjuntamente 85,7% da variância total (PC1 = 54,8%, PC2 = 30,6%), indicando que grande parte da variabilidade foi representada em apenas duas dimensões.

A Dimensão 1 foi fortemente influenciada por Major, Minor, Area e Perim, sugerindo que este eixo reflete principalmente diferenças relacionadas ao tamanho das sementes. Já a Dimensão 2 apresentou maior correlação com AR e Round, indicando que este componente captura variações associadas à forma das sementes. O rendimento (Rend) apresentou baixa contribuição em ambas as dimensões, sugerindo que este caráter não foi determinante na separação dos genótipos na PCA.

A projeção dos acessos no plano da PCA, combinada com o agrupamento k-means aplicado aos valores de BLUPs, revelou três agrupamentos fenotípicos distintos. Os grupos diferiram em relação às características morfológicas das sementes, com o grupo 2 apresentando acessos com sementes maiores, enquanto os outros grupos incluíram acessos com sementes menores, mais redondas e com maiores rendimentos. Essa separação evidencia a heterogeneidade fenotípica do BAG e a possibilidade de selecionar genótipos divergentes para cruzamentos direcionados.

Estudos conduzidos com mamoneira e outras oleaginosas confirmam a relevância da integração entre análise de componentes principais (PCA) e fenotipagem digital. Vilar *et al.* (2015) aplicaram a PCA aos dados de cor obtidos por imagens digitais de sementes de mamoneira, identificando agrupamentos consistentes em uma análise exploratória. Já Kim *et al.* (2022) analisaram mais de 53 mil sementes de soja por imagens digitais e demonstraram a eficácia da PCA na identificação de genótipos divergentes, enquanto Khatun *et al.* (2023) destacaram a fenotipagem digital como ferramenta eficiente para seleção de genitores em programas de melhoramento de soja.

Nesse contexto, a PCA aplicada a variáveis morfológicas extraídas de imagens digitais fornece uma abordagem rápida e precisa para explorar a variabilidade genética, reduzir redundâncias no BAG e orientar a escolha de genitores divergentes, aumentando a eficiência dos programas de melhoramento em mamoneira Kim *et al.* (2025).

Em síntese, estudos que estimem parâmetros genéticos, realizem a predição do peso de cem sementes e investiguem a diversidade com base em descritores morfológicos extraídos de imagens digitais de sementes de mamoneira ao longo de diferentes anos agrícolas ainda são escassos. Os resultados obtidos neste trabalho evidenciam a originalidade e a relevância da pesquisa, ao fornecer subsídios para

estratégias de melhoramento genético mais precisas e eficientes no NBIO, a partir da incorporação desses descritores. Perspectivas futuras incluem a integração das análises fenotípicas digitais com ferramentas genômicas, como a seleção genômica e os estudos de associação ampla do genoma (GWAS), o que representa um caminho promissor para identificar regiões genômicas relacionadas a caracteres de interesse e acelerar o ganho genético em mamoneira.

5. CONCLUSÕES

A fenotipagem por visão computacional se mostrou uma abordagem viável e precisa para mensuração das características morfológicas de sementes de mamoneira, oferecendo uma alternativa eficiente ao método tradicional.

As características extraídas por imagem permitiram treinar e validar modelos de aprendizado de máquina para estimativa do peso de 100 sementes. O algoritmo Ridge Regression (RR) apresentou melhor desempenho ($R^2 = 0,88$; RMSE = 3,83; MAE = 3,19), seguido por K-Nearest Neighbors (KNN) e Deep Neural Network (DNN), que também apresentaram resultados satisfatórios.

As variáveis extraídas por imagem capturaram 85,7% da variância total na PCA, revelando três grupos fenotípicos distintos no BAG de mamoneira. Características como Major ($H^2 = 0,88$) e AR ($H^2 = 0,87$) apresentaram alta herdabilidade, enquanto Round ($H^2 = 0,79$), Perímetro ($H^2 = 0,72$) e Área ($H^2 = 0,67$) foram classificadas com herdabilidade média. Todas essas características exibiram menor variância do erro e boa capacidade de refletir a variabilidade genética entre os acessos, mostrando-se confiáveis para a avaliação e a seleção nas fases iniciais de programas de melhoramento.

6. REFERÊNCIAS

- ABEKOON, T.; SAJINDRA, H.; RATHNAYAKE, N.; EKANAYAKE, I. U.; JAYAKODY, A.; RATHNAYAKE, U. A novel application with explainable machine learning (SHAP and LIME) to predict soil N, P, and K nutrient content in cabbage cultivation. **Smart Agricultural Technology**, v. 11, p. 100879, 2025. DOI: <https://doi.org/10.1016/j.atech.2025.100879>
- ACOSTA-NAVARRETE, M. S.; TIGREROS, J. A.; HOLGADO-APAIZA, L. A.; FLORES-BALDERAS, J. N.; ADAME, J. M. B. Morphometric, weight, viability, and germination analysis of castor bean seeds (*Ricinus communis*) under two temperature and relative humidity conditions. **Agro Productividad**, 2023. DOI: <https://doi.org/10.32854/agrop.v15i4.2484>
- BAEK, J.; LEE, E.; KIM, N.; KIM, S. L.; CHOI, I.; JI, H.; CHUNG, Y. S.; CHOI, M. S.; MOON, J. K.; KIM, K. H. High throughput phenotyping for various traits on soybean seeds using image analysis. **Sensors**, v. 20, n. 1, p. 248, 2020. DOI: <https://doi.org/10.3390/s20010248>
- BLAND, J. M.; ALTMAN, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. **The lancet**, v. 327, n. 8476, p. 307-310, 1986. DOI: [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)
- CAPINAN, G. C. S.; SILVA, S. A.; SOUZA, D. R. D.; SANTOS, L. A. D.; JESUS, I. B. S. D. VARIABILITY OF F 2 PROGENIES OF CASTOR BEAN BY MEANS OF MORPHOAGRONOMIC DESCRIPTORS. **Revista Caatinga**, v. 35, p. 53-67, 2022. DOI: <https://doi.org/10.1590/1983-21252022v35n106rc>
- CHEN, B.; LI, J.; YAO, S.; WANG, G.; WANG, X. Seed-specific expression of phosphatidate phosphohydrolases increases soybean oil content and seed weight. **Biotechnology for Biofuels and Bioproducts**, v. 18, n. 1, p. 23, 2025. DOI: <https://doi.org/10.1186/s13068-025-02620-x>
- COUTO, D. P.; OLIVEIRA, W. B. D. S.; OLIVEIRA, J. S.; GUILHEN, J. H. S.; BERNARDES, C. D. O.; POSSE, S. C. P.; FERREIRA, M. F. D. S.; FERREIRA, A. Analysis of the effect of the interaction of genotype and environment on the yield stability of maize varieties; genetic resources for breeding. **Agronomy**, v. 13, n. 8, p. 1970, 2023. DOI: <https://doi.org/10.3390/agronomy13081970>
- DIAO, C.; ZHUO, Y.; MAO, R.; LI, W.; DU, H.; ZHOU, L.; LIU, J. Weighted Kernel Ridge Regression to Improve Genomic Prediction. **Agriculture**, v. 15, n. 5, p. 445, 2025. DOI: <https://doi.org/10.3390/agriculture15050445>
- DELVADIYA, I. R.; MADARIYA, R. B.; GINOYA, A. V.; PATEL, J. Uncovering the inheritance mechanisms of yield and oil content in castor (*Ricinus communis* L.): a 21-generation study. **Euphytica**, v. 220, n. 2, p. 20, 2024. DOI: <https://doi.org/10.1007/s10681-023-03277-z>
- DUC, N. T.; RAMLAL, A.; RAJENDRAN, A.; RAJU, D.; LAL, S.K.; KUMAR, S.; SAHOO, R. N.; CHINNUSAMY, V. Image-based phenotyping of seed architectural traits and prediction of seed weight using machine learning models in soybean.

Frontiers in Plant Science, v. 14, p. 1206357, 2023. DOI: <https://doi.org/10.3389/fpls.2023.1206357>

ELHAM, R. S. S.; SOLIMAN, M. S. A. Diversity Assessment by Molecular Barcoding and Seed Morphology in *Ricinus communis* L. **Baghdad Science Journal**, v. 18, n. 1, p. 41, 2021. DOI: [https://doi.org/10.21123/bsj.2021.18.1\(Suppl.\).0708](https://doi.org/10.21123/bsj.2021.18.1(Suppl.).0708)

EL-HENDAWY, S.; TAHIR, M. U.; AL-SUHAIBANI, N.; ELSAYED, S.; ELSHERBINY, O.; ELSHARAWY, H. Potential of thermal and RGB imaging combined with artificial neural networks for assessing salt tolerance of wheat genotypes grown in real-field conditions. **Agronomy**, v. 14, n. 7, p. 1390, 2024. DOI: <https://doi.org/10.3390/agronomy14071390>

FIGÀS, M. R.; PROHENS, J.; CASANOVA, C.; FERNÁNDEZ-DE-CÓRDOVA, P.; SOLER, S. Variation of morphological descriptors for the evaluation of tomato germplasm and their stability across different growing conditions. **Scientia Horticulturae**, v. 238, p. 107-115, 2018. DOI: <https://doi.org/10.1016/j.scienta.2018.04.039>

FOX, J.; WEISBERG, S. **An R Companion to Applied Regression**, Third edition. Sage, Thousand Oaks CA. <https://www.john-fox.ca/Companion/>. 2018.

FRÍAS, M. P.; MARTÍNEZ, F. An Ensemble for Automatic Time Series Forecasting with K-nearest Neighbors. **IEEE Access**, 2025. DOI: [10.1109/ACCESS.2025.3525561](https://doi.org/10.1109/ACCESS.2025.3525561)

GÓMEZ, J. J. M.; SAADAOU, E.; CERVANTES, E. Seed shape of castor bean (*Ricinus communis* L.) grown in different regions of Tunisia. **Journal of Agriculture and Ecology Research International**, v. 8, n. 1, 2016. DOI: [10.9734/JAERI/2016/23934](https://doi.org/10.9734/JAERI/2016/23934)

ISAZA, C.; ANAYA, K.; DE PAZ, J. Z.; VASCO-LEAL, J. F.; HERNANDEZ-RIOS, I.; MOSQUERA-ARTAMONOV, J. D. Image analysis and data mining techniques for classification of morphological and color features for seeds of the wild castor oil plant (*Ricinus communis* L.). **Multimedia Tools and Applications**, v. 77, n. 2, p. 2593-2610, 2018. DOI: <https://doi.org/10.1007/s11042-017-4438-y>

JEWAN, S. Y. Y.; PAGAY, V.; BILLA, L.; TYERMAN, S. D.; GAUTAM, D.; SPARKES, D.; CHAI, H. H.; SINGH, A. The feasibility of using a low-cost near-infrared, sensitive, consumer-grade digital camera mounted on a commercial UAV to assess *Bambara groundnut* yield. **International Journal of Remote Sensing**, v. 43, n. 2, p. 393-423, 2022. DOI: <https://doi.org/10.1080/01431161.2021.1974116>

JIANG, M.; GAO, X.; DANG, Q.; RUAN, J. An evolutionary multitasking algorithm based on k-nearest neighbors pre-selection strategy for constrained multi-objective optimization. **Expert Systems with Applications**, p. 127768, 2025. DOI: <https://doi.org/10.1016/j.eswa.2025.127768>

KARATZOGLOU, A.; SMOLA, A.; HORNIK, K.; ZEILEIS, A. kernlab-an S4 package for kernel methods in R. **Journal of statistical software**, v. 11, p. 1-20, 2004. DOI: <https://doi.org/10.18637/jss.v011.kernlab>

10.18637/jss.v011.i09

KASSAMBARA, A.; MUNDT, F. Package 'factoextra'. **Extract and visualize the results of multivariate data analyses**, v. 76, n. 2, p. 10.18637, 2017.

KASSAMBRA, A. **_rstatix: pipe-friendly framework for basic statistical tests_**. **R package version 0.7.1**. <https://cran.r-project.org/package=rstatix>. 2022.

KAUSHAL, S.; GILL, H. S.; BILLAH, M. M.; KHAN, S. N.; HALDER, J.; BERNARDO, A.; AMAND, P. S.; BAI, G.; GLOVER, K.; MAIMAITIJIANG, M.; SEHGAL, S. K. Enhancing the potential of phenomic and genomic prediction in winter wheat breeding using high-throughput phenotyping and deep learning. **Frontiers in Plant Science**, v. 15, p. 1410249, 2024. DOI: <https://doi.org/10.3389/fpls.2024.1410249>

KHATUN, M.; ISLAM, M. R.; KHAN, M. A. R.; HOSSAIN, M. K.; ISLAM, A. K. M. A. Principal component analysis and estimated breeding values for selecting suitable parental genotypes in rice (*Oryza sativa* L.). **Sabao Journal of Breeding & Genetics**, v. 55, n. 6, 2023. DOI: <http://doi.org/10.54910/sabao2023.55.6.3>

KIM, J. M.; LEE, J. W.; KIM, D. J.; LYU, J. I.; BAEK, J.; HA, B. K.; KWON, S. J. Image-based GWAS identifies the genetic architecture of seed-related traits in a soybean mutant population. **Molecular Breeding**, v. 45, n. 8, p. 67, 2025. DOI: <https://doi.org/10.1007/s11032-025-01584-y>

KIM, S. H.; JO, J. W.; WANG, X.; SHIN, M. J.; HUR, O. S.; HA, B. K.; HAHN, B. S. Diversity characterization of soybean germplasm seeds using image analysis. **Agronomy**, v. 12, n. 5, p. 1004, 2022. DOI: <https://doi.org/10.3390/agronomy12051004>

KIM, J. Y.; CHUNG, Y. S. A short review of RGB sensor applications for accessible high-throughput phenotyping. **Journal of Crop Science and Biotechnology**, p. 1-5, 2021. DOI: <https://doi.org/10.1007/s12892-021-00104-6>

KUHN, M.; WING, J.; WESTON, S.; WILLIAMS, A.; KEEFER, C.; ENGELHARDT, A.; COOPER, T.; MAYER, Z.; KENKEL, B.; Team, R. C. Package 'caret'. **The R Journal**, v. 223, n. 7, p. 48, 2020.

KUZNETSOVA, A.; BROCKHOFF, P. B.; CHRISTENSEN, R. H. B. lmerTest package: tests in linear mixed effects models. **Journal of statistical software**, v. 82, p. 1-26, 2017. DOI: 10.18637/jss.v082.i13

LANDONI, M.; BERTAGNON, G.; GHIDOLI, M.; CASSANI, E.; ADANI, F.; PILU, R. Opportunities and challenges of castor bean (*Ricinus communis* L.) genetic improvement. **Agronomy**, v. 13, n. 8, p. 2076, 2023. DOI: <https://doi.org/10.3390/agronomy13082076>

LÊ, S.; JOSSE, J.; HUSSON, F. "FactoMineR: A Package for Multivariate Analysis." **Journal of Statistical Software**, v. 25, n. 1, p. 1–18, 2008. DOI: <https://doi.org/10.18637/jss.v025.i01>

- MARQUES, R. F.; GOMES, F. R.; MARTINS, C. C.; MARCHI, S. R. D.; MARTINS, D. Field emergence of *Marandu palisadegrass* influenced by specific seed weight and sowing depth. **Revista Ciência Agronômica**, v. 56, p. e202392211, 2024. DOI: <https://doi.org/10.5935/1806-6690.20250027>
- MIRANDA, M. C. C.; AONO, A. H.; PINHEIRO, J. B. A novel image-based approach for soybean seed phenotyping using machine learning techniques. **Crop Science**, v. 63, n. 5, p. 2665-2684, 2023. DOI: <https://doi.org/10.1002/csc2.21032>
- PARVEEN, N.; UMER, S.; TAN, C.; JABBAR, A.; KANWAL, B.; HAIDER, I.; RAZA, W.; USMA, A.; MEHMOOD, A.; JUNAID, M. B.; ABBASI, S. H.; ALFAGHAM, A. T.; IQBAL, R. Multivariate and Association Analyses of Various Seed Yield Contributing Traits Divulge Genetic Diversity Among *Pisum sativum* L. Genotypes. **Plant Molecular Biology Reporter**, p. 1-9, 2025. DOI: <https://doi.org/10.1007/s11105-025-01533-1>
- PATEL, J.; PATEL, S.; COOK, L.; FALLEN, B. D.; KOEBERNICK, J. Soybean genome-wide association study of seed weight, protein, and oil content in the southeastern USA. **Molecular Genetics and Genomics**, v. 300, n. 1, p. 1-14, 2025. DOI: <https://doi.org/10.1007/s00438-025-02228-8>
- PLUTENKO, L.; RADCHUK, V.; MAYER, S.; KEIL, P.; ORTLEB, S.; WAGNER, S.; LEHMANN, V.; ROLLETSCHEK, H.; BORISJUK, L. MRI-Seed-Wizard: combining deep learning algorithms with magnetic resonance imaging enables advanced seed phenotyping. **Journal of Experimental Botany**, v. 76, n. 2, p. 393-410, 2025. DOI: <https://doi.org/10.1093/jxb/erae408>
- R Core Team. R: A language and environment for statistical computing. **Version 4.4.1, R Foundation for Statistical Computing, Vienna, Austria**, 2023. Disponível em: <https://www.R-project.org/>. Acesso em: 28 de outubro de 2024.
- REVELLE, W.; REVELLE, M. W. Package 'psych'. **The comprehensive R archive network**, v. 337, n. 338, p. 161-165, 2015.
- REZZOUK, F. Z.; GRACIA-ROMERO, A.; KEFAUVER, S. C.; GUTIÉRREZ, N. A.; ARANJUELO, I.; SERRET, M. D.; ARAUS, J. L. Remote sensing techniques and stable isotopes as phenotyping tools to assess wheat yield performance: Effects of growing temperature and vernalization. **Plant Science**, v. 295, p. 110281, 2020. DOI: <https://doi.org/10.1016/j.plantsci.2019.110281>
- RIBEIRO, L. P.; SANTOS, D. M. B.; NETO, L. A. L.; NETO, L. R. S.; BARBOSA, M. F.; CUNHA, T. J. F. Levantamento detalhado dos solos, capacidade de uso, e classificação de terras para irrigação da estação de plasticultura da Universidade Federal da Bahia/Politeo em Cruz das Almas (BA). **Revista Brasileira de Ciência do Solo, Campinas**, v. 19, n. 1, p. 105-113, 1995.
- RINKER, T. W.; KURKIEWICZ, D. Pacman: package management for R. **Buffalo, New York**. <http://github.com/trinker/pacman>. 2018.
- ROKA, P.; Shrestha, S.; ADHIKARI, S. P.; NEUPANE, A.; SHREEPAILI, B.; BISTA,

M. K. A review on genetic parameters estimation, trait association, and multivariate analysis for crop improvement. **Archives of Agriculture and Environmental Science**, v. 9, n. 3, p. 618-625, 2024. DOI: <https://doi.org/10.26832/24566632.2024.0903029>

SILVA, A. R.; SILVA, A. S.; SANTOS L. A.; SOUZA, D. R.; ARAUJO, G. M.; DANTAS, J. L. L.; LEITE, E. D.; DANTAS, A. C. V. L. Characterization and performance of castor bean lineages and parents at the UFRB germplasm bank. **PLoS ONE**, v. 14, n. 1, p. e0209335, 2019. DOI: <https://doi.org/10.1371/journal.pone.0209335>

SINGH, S. K.; VIDYARTHI, S. K.; TIWARI, R. Machine learnt image processing to predict weight and size of rice kernels. **Journal of Food Engineering**, v. 274, p. 109828, 2020. DOI: <https://doi.org/10.1016/j.jfoodeng.2019.109828>

SIWALE, J.; LABUSCHAGNE, M.; GERRANO, A. S.; MBUMA, N. W. Phenotypic diversity and characterization of the Southern African Bambara groundnut Germplasm collection for grain yield and yield components. **Agronomy**, v.12, n.8, p. 1811, 2022. DOI: <https://doi.org/10.3390/agronomy12081811>

SOLIMAN, E. R. S.; SOLIMAN, M. S. A. Diversity Assessment by Molecular Barcoding and Seed Morphology in *Ricinus communis* L. **Baghdad Science Journal**, v. 18, n. 1, p. 41, 2021. DOI: [https://doi.org/10.21123/bsj.2021.18.1\(Suppl.\).0708](https://doi.org/10.21123/bsj.2021.18.1(Suppl.).0708)

SUNOJ, S.; IGATHINATHANE, C.; FLORES, J. P.; SIDHU, H.; MONONO, E.; SCHATZ, B.; ARCHER, D.; HENDRICKSON, J. Crop row identification and plant cluster segmentation for stand count from UAS imagery based on profile and geometry. **Smart Agricultural Technology**, v. 11, p. 100938, 2025. DOI: <https://doi.org/10.1016/j.atech.2025.100938>

VILAR, W. T.; ARANHA, R. M.; MEDEIROS, E. P.; PONTES, M. J. Classification of individual castor seeds using digital imaging and multivariate analysis. **Journal of the Brazilian Chemical Society**, v. 26, n. 1, p. 102-109, 2015. DOI: <https://doi.org/10.5935/0103-5053.20140221>

VYMYSLICKÝ, T.; TRNĚNÝ, O.; RIETMAN, H.; BALKO, C.; ĐORĐEVIĆ, V.; RANĐELOVIĆ, P.; DYBOVÁ, M. Phenotypic characterization of soybean genetic resources at multiple locations: breeding implications for enhancing environmental resilience, yield and protein content. **Frontiers in Plant Science**, v. 16, p. 1422162, 2025. DOI: <https://doi.org/10.3389/fpls.2025.1422162>

WITT, T. W.; FLYNN, K. C.; ZOZ, T.; LEE, T. O.; MONTEIRO, J. E. A site suitability analysis for castor (*Ricinus communis* L.) production during Brazil's second harvest incorporating disease prediction. **Heliyon**, v. 9, n. 8, 2023. DOI: [10.1016/j.heliyon.2023.e18981](https://doi.org/10.1016/j.heliyon.2023.e18981)

YANG, S.; ZHENG, L.; HE, P.; WU, T.; SUN, S.; WANG, M. High-throughput soybean seeds phenotyping with convolutional neural networks and transfer learning. **Plant Methods**, v. 17, n. 1, p. 50, 2021. DOI: <https://doi.org/10.1186/s13007-021->

00749-y

ZHANG, H.; JI, J.; MA, H.; GUO, H.; LIU, N.; CUI, H. Wheat seed phenotype detection device and its application. **Agriculture**, v. 13, n. 3, p. 706, 2023. DOI: <https://doi.org/10.3390/agriculture13030706>

ZHOU, M. Genotype and Environmental Effects on Sugarcane Flowering in South Africa. **Sugar Tech**, v. 24, n. 2, p. 430-437, 2022. DOI: <https://doi.org/10.1007/s12355-021-01025-y>

CAPÍTULO II

CARACTERIZAÇÃO QUANTITATIVA DA COLORAÇÃO DE ESTRUTURAS VEGETAIS DE MAMONEIRA POR VISÃO COMPUTACIONAL

CARACTERIZAÇÃO QUANTITATIVA DA COLORAÇÃO DE ESTRUTURAS VEGETAIS DE MAMONEIRA POR VISÃO COMPUTACIONAL

RESUMO: A caracterização quantitativa da cor por visão computacional representa uma estratégia objetiva e eficaz para reduzir a subjetividade das avaliações visuais e ampliar o uso de descritores morfológicos em programas de melhoramento genético. Este estudo teve como objetivo caracterizar quantitativamente a coloração de folhas, galhos, pecíolos, frutos e sementes de germoplasma de mamoneira por meio de visão computacional, com aplicação em programas de melhoramento genético. As estruturas foram avaliadas visualmente e fotografadas em miniestúdio com câmera RGB sob iluminação controlada, com pré-processamento no ImageJ® e análise no software R para extração de histogramas no espaço RGB e variáveis colorimétricas no espaço CIELAB (L^* , a^* , b^* , Hue^* e $Chroma^*$). Adicionalmente, componentes de variância e herdabilidade no sentido amplo foram estimados a partir de modelos lineares mistos via REML/BLUPs. A avaliação visual classificou as linhagens em três grupos de coloração para galhos e pecíolos, quatro para folhas e frutos e seis para sementes. A análise hierárquica baseada na métrica EMD evidenciou contrastes cromáticos consistentes, destacando a separação de galhos vermelhos (38%), verde-claros (20%) e verde-avermelhados (42%), bem como a distinção de pecíolos vermelhos (40%) em relação aos verdes. Para sementes, seis classes foram identificadas, com predominância de coloração preta (65%), seguida por padrões marrons e acinzentados. No espaço CIELAB, as variáveis colorimétricas permitiram identificar padrões específicos de coloração, reforçando a precisão do modelo em discriminar classes colorimétricas das estruturas. Além disso, variâncias genéticas significativas ($p < 0,05$ a $p < 0,0001$) foram observadas para as variáveis colorimétricas de todas as estruturas, exceto folhas, que apresentaram baixa herdabilidade ($H^2 = 0,23-0,56$), indicando maior influência ambiental. Em contraste, galhos (L^* , a^* , b^* , $Hue^* > 0,90$), pecíolos (L^* , a^* , b^* , $Hue^* > 0,89$) e sementes exibiram herdabilidades elevadas, com destaque para Hue^* em sementes ($H^2 = 1,00$), confirmando forte controle genético da coloração para essas estruturas. De forma integrada, a combinação das análises nos espaços RGB e CIELAB mostrou-se complementar às avaliações visuais, permitindo identificar contrastes cromáticos consistentes entre linhagens e estruturas vegetais. As elevadas estimativas de herdabilidade e a variância genética evidenciam o potencial das variáveis de cor como marcadores fenotípicos robustos, úteis tanto na seleção de linhagens de mamoneira quanto na manutenção da diversidade cromática em bancos de germoplasma.

Palavras-chave: CIELAB, Fenotipagem digital, Herdabilidade, Melhoramento genético, RGB, Variabilidade cromática.

QUANTITATIVE CHARACTERIZATION OF CASTOR BEAN PLANT STRUCTURES' COLORATION USING COMPUTER VISION

ABSTRACT: Quantitative color characterization using computer vision represents an objective and effective strategy to reduce the subjectivity of visual assessments and expand the use of morphological descriptors in plant breeding programs. This study aimed to quantitatively characterize the color of leaves, branches, petioles, fruits, and seeds of castor bean germplasm through computer vision, with applications in genetic improvement programs. The plant structures were visually evaluated and photographed in a mini-studio using an RGB camera under controlled lighting conditions. Preprocessing was performed in ImageJ®, and subsequent analyses were conducted in R software to extract histograms in the RGB color space and colorimetric variables in the CIELAB space (L^* , a^* , b^* , Hue^* , and $Chroma^*$). In addition, variance components and broad-sense heritability were estimated using mixed linear models via REML/BLUPs. Visual evaluation classified the lines into three color groups for branches and petioles, four for leaves and fruits, and six for seeds. Hierarchical clustering based on the EMD metric revealed consistent chromatic contrasts, highlighting the separation of red branches (38%), light-green (20%), and reddish-green (42%), as well as the distinction between red (40%) and green petioles. For seeds, six color classes were identified, with a predominance of black seeds (65%), followed by brown and grayish patterns. In the CIELAB space, colorimetric variables allowed the identification of specific color patterns, reinforcing the model's accuracy in discriminating colorimetric classes of the plant structures. Moreover, significant genetic variances ($p < 0.05$ to $p < 0.0001$) were observed for the colorimetric variables of all structures except leaves, which exhibited low heritability ($H^2 = 0.23\text{--}0.56$), indicating a greater environmental influence. In contrast, branches (L^* , a^* , b^* , $Hue^* > 0.90$), petioles (L^* , a^* , b^* , $Hue^* > 0.89$), and seeds showed high heritability, with Hue^* in seeds ($H^2 = 1.00$) confirming strong genetic control of color for these structures. Overall, the combination of analyses in RGB and CIELAB color spaces proved complementary to visual evaluations, enabling the identification of consistent chromatic contrasts among lines and plant structures. The high heritability estimates and significant genetic variance demonstrate the potential of color variables as robust phenotypic markers, useful both for selecting castor bean lines and for maintaining chromatic diversity in germplasm banks.

Keywords: CIELAB, Digital phenotyping, Heritability, Plant breeding, RGB, Chromatic variability.

1. INTRODUÇÃO

A mamoneira (*Ricinus communis* L.) apresenta ampla variabilidade intraespecífica, especialmente na coloração de estruturas como galho, folha, pecíolo, fruto e semente (Nour *et al.*, 2023). Essa diversidade fenotípica constitui um critério relevante em programas de melhoramento genético, uma vez que a cor é normalmente controlada por poucos genes, tornando-as marcadores morfológicos eficazes para análise genética, contribuindo para a conservação e o uso eficiente da variabilidade genética disponível (Wahibah *et al.*, 2020).

A categorização de cores com base apenas em avaliações visuais, entretanto, apresenta limitações relacionadas à precisão, objetividade e reprodutibilidade. A percepção humana da cor é influenciada por fatores fisiológicos, como diferenças genéticas na sensibilidade das células fotorreceptoras da retina e processos de adaptação neural, além de fatores ambientais, como variações na iluminação, o que pode resultar em inconsistências e comprometer a padronização dos dados (Du *et al.*, 2021; Bosten, 2022; Reynolds; Singh, 2024; Karimipour; Witzel, 2024).

Nesse contexto, a análise quantitativa da cor por meio da visão computacional configura-se como uma abordagem promissora, pois possibilita a geração de dados reprodutíveis e amplia a capacidade de caracterização morfológica, permitindo maior objetividade na detecção de variações cromáticas (Loarca *et al.*, 2024).

A análise quantitativa de cor tem sido amplamente empregada em diferentes estudos de germoplasma, incluindo estimativas indiretas de pigmentos em flores de acessos de crisântemo e híbridos de lírio, avaliação da coloração de grãos em genótipos de trigo, de frutos em genótipos de goiaba, bem como de folhas e vagens em acessos de feijão (Lu *et al.*, 2021; Bajgain; Li; Anderson, 2022; Li *et al.*, 2022; Celebioglu *et al.*, 2024; Mathiazhagan *et al.*, 2024; Wan *et al.*, 2024).

Os estudos anteriores de análise quantitativa de cor em mamoneira têm se concentrado principalmente nas sementes, enquanto outras estruturas permanecem pouco exploradas (Vilar *et al.*, 2015; Gómez; Saadaqui; Cervantes, 2016; Isaza *et al.*, 2018). Como folhas, galhos, pecíolos e frutos também apresentam variação cromática consistente, a ampliação dessa abordagem pode fornecer descritores adicionais para o melhoramento genético.

Diante disso, o objetivo deste trabalho foi caracterizar quantitativamente a coloração de folha, galho, pecíolo, fruto e semente de germoplasma de mamoneira

por meio de visão computacional para aplicação em programas de melhoramento genético.

2. MATERIAL E MÉTODOS

2.1. Área experimental

O experimento foi conduzido na área experimental do Núcleo de Melhoramento Genético e Biotecnologia (NBIO), vinculado ao Centro de Ciências Agrárias, Ambientais e Biológicas (CCAAB) da Universidade Federal do Recôncavo da Bahia (UFRB), localizada no município de Cruz das Almas, Bahia (13°12'40"S; 39°24'23"W; 220 m de altitude) (Ribeiro *et al.*, 1995).

O preparo do solo consistiu em aração seguida de duas gradagens. Após análise química do solo, foi realizada adubação de fundação com 15 kg ha⁻¹ de nitrogênio (N), 50 kg ha⁻¹ de fósforo (P) e 30 kg ha⁻¹ de potássio (K), aplicados diretamente nas covas (15 cm de profundidade × 20 cm de diâmetro). Adicionalmente, 20 kg ha⁻¹ de N foram aplicados em cobertura.

As mudas foram produzidas em casa de vegetação, utilizando sacos de polietileno de 1,5 L, contendo substrato composto por solo argiloso e composto orgânico na proporção de 2:1 (v/v) peneirados. Devido ao baixo vigor germinativo das sementes, foi adotado o uso de mudas em substituição à semeadura no campo. Essa estratégia permitiu maior controle das condições de germinação e do desenvolvimento inicial, favorecendo a obtenção de plântulas vigorosas e aptas ao transplante.

A semeadura foi realizada em 23 de novembro de 2023 e o transplante em 29 de janeiro de 2024, com espaçamento de 3 × 1 m e duas aplicações de fungicida com princípio ativo Tiofanato-metílico e dosagem 60g 100L⁻¹ para controle de mofo cinzento.

2.2. Delineamento experimental

Adotou-se o delineamento em blocos aumentados, com quatro blocos, para a condução do experimento. As testemunhas foram compostas pelas cultivares BRS Energia, BRS Paraguaçu e IAC 2028, além das variedades Tomy, Couro e NOA, sendo todas repetidas nos blocos. Os tratamentos regulares foram representados por 65 linhagens endogâmicas do Banco Ativo de Germoplasma (BAG) de mamoneira, distribuídos aleatoriamente dentro dos blocos.

O experimento foi composto por oito linhas de plantio com 50 plantas por linha. As linhas laterais, bem como a primeira e última planta de cada linha, foram utilizadas como bordaduras.

2.3. Material vegetal

Foram avaliadas linhagens pertencentes ao BAG de mamoneira do NBIO-UFRB, desenvolvido a partir de hibridações controladas entre cinco genitores: BRS Nordestina, BRS Paraguaçu, EBDA MPA 34, Mirante 10 e Sipeal 28. As populações F1 foram autofecundadas para obtenção da geração F2, conduzida até as gerações F5/F6 pelo método Single Seed Descent (SSD), onde foi realizada a seleção das linhagens. Após mais um ciclo de autofecundação, a geração F7 foi obtida e utilizada no experimento.

2.4. Categorização visual

A categorização visual direta foi baseada nas cores do galho, pecíolo, folha, fruto e semente de 65 linhagens isentas de sintomas de pragas e doenças e sem sinais de deficiências nutricionais. As estruturas foram coletadas da porção mediana da planta aos nove meses após a semeadura. Essa classificação foi realizada conforme as classes fenotípicas de cores descritas no Formulário de Instruções para Execução dos Ensaio de Distingibilidade, Homogeneidade e Estabilidade de Cultivares de Mamona (*Ricinus communis* L.) do Ministério da Agricultura, Pecuária e Abastecimento – MAPA (BRASIL, 2008).

2.5. Aquisição e processamento de imagens digitais

As imagens do galho, pecíolo, fruto e folha foram capturadas utilizando um iPhone 11 (lente angular 26 mm, sensor RGB), utilizando os seguintes parâmetros: abertura f/1.8 e ISO 32. Para padronização das condições de iluminação, foi utilizado um miniestúdio fotográfico com dimensões de 40 × 40 cm, equipado com uma barra de LED e uma abertura circular na parte superior. As amostras foram centralizadas sobre um fundo branco, acompanhadas por uma etiqueta de identificação da respectiva linhagem.

Já as imagens das sementes foram capturadas com uma câmera Canon EOS REBEL T5i (lente 18–55 mm, sensor RGB), utilizando os seguintes parâmetros: velocidade do obturador 1/80s, abertura f/5.0, ISO 200 e balanço de branco

personalizado. Para padronização da iluminação das imagens, foi utilizado um mini estúdio fotográfico customizado, composto por suporte retangular com câmera acoplada na parte superior e duas fontes de luz LED ajustáveis nas laterais. As sementes foram dispostas sobre fundo azul, acompanhadas por uma identificação do genótipo e uma régua.

As imagens capturadas no formato JPEG foram submetidas a um pré-processamento visando garantir maior uniformidade e qualidade para as análises de cor. O recorte das imagens foi realizado no software ImageJ® (versão 1.54g), a fim de eliminar regiões externas irrelevantes e delimitar exclusivamente as estruturas de interesse. Em seguida, o aplicativo Photoroom® foi utilizado para a remoção do fundo branco, permitindo a exclusão de ruídos visuais e favorecendo a extração precisa dos valores RGB (Red, Green, Blue).

2.6. Extração e análise do perfil quantitativo de cores

Após o pré-processamento, as imagens foram importadas para o software R (versão 4.4.1) (R Core Team, 2023), onde foram convertidas em matrizes de pixels no espaço de cor RGB. O processamento foi realizado com funções do pacote *colordistance* (Weller; Westneat, 2019), que possibilita a extração e comparação de padrões de cores em imagens digitais. Inicialmente, as imagens foram carregadas e submetidas a filtros para considerar apenas áreas com valores elevados de intensidade nos três componentes de cor, removendo ruídos residuais de fundo (limites definidos com $\text{lower} = c(0.99, 0.99, 0.99)$ e $\text{upper} = c(1, 1, 1)$). Em seguida, foram gerados histogramas RGB com quatro intervalos (*bins*) por canal para cada imagem.

Para cada estrutura, foi obtida uma matriz de distâncias de cor entre as imagens utilizando a métrica Earth Mover's Distance (EMD). Essa métrica quantifica o esforço necessário para transformar uma distribuição de cores, considerando simultaneamente as diferenças de tonalidade e a proporção relativa das cores (Weller; Westneat, 2019). A matriz de distâncias obtida foi empregada na análise de agrupamento hierárquico, com o objetivo de identificar padrões de similaridade colorimétrica entre as linhagens. O agrupamento foi realizado pelo método de ligação completa (*complete linkage*), utilizando o EMD como medida de dissimilaridade. As análises foram conduzidas com os pacotes *ape* (Paradis; Schliep, 2019) e *ggtree* (Yu, 2020) para a construção e visualização dos dendrogramas, além

do circlize (Gu *et al.*, 2014) para a representação em formato circular.

2.7. Obtenção de variáveis de cor derivadas

A análise quantitativa da cor foi realizada a partir de imagens digitais para obtenção das coordenadas L^* , a^* , b^* , Hue e Chroma. O processamento das imagens e a extração dos valores de cor foram realizados no ambiente estatístico R versão 4.4.1 (R Core Team, 2023), com o auxílio dos pacotes *colorspace* (Zeileis; Hornik; Murrell, 2020), *png* (Urbanek, 2022), *jpeg* (Urbanek, 2025) e *dplyr* (Wickham *et al.*, 2023).

Cada imagem foi lida e convertida do espaço de cor RGB para o espaço de cor CIELAB (L^* , a^* , b^*), que representa a cor de maneira perceptual uniforme. A conversão foi implementada por meio do pacote *colorspace*, onde o canal L^* representa a luminosidade (0 = preto, 100 = branco), o canal a^* varia de verde ($-a^*$) a vermelho ($+a^*$), e o canal b^* varia de azul ($-b^*$) a amarelo ($+b^*$).

Um filtro foi aplicado para remover o fundo branco das imagens e isolar as amostras de interesse. O filtro foi baseado nos valores L^* , a^* e b^* de cada pixel. Pixels com um valor de L^* maior que 90 e valores de a^* e b^* próximos de zero (neste caso, com um valor absoluto menor que 5) foram identificados como parte do fundo branco e, portanto, excluídos da análise, garantindo que apenas os pixels da amostra fossem considerados.

Após a filtragem, os valores dos canais L^* , a^* e b^* foram extraídos de cada pixel remanescente. A mediana de cada canal (L^* , a^* e b^*) foi calculada para cada imagem, a fim de obter um valor representativo que fosse menos suscetível a valores atípicos (outliers) do que a média.

Com base nos valores medianos de a^* e b^* , o Croma (Chroma*) e o Ângulo de Tonalidade (Hue*) foram calculados para cada imagem, utilizando as seguintes equações:

$$Hue = \frac{180}{\pi} \times \cos \left[a / \sqrt{(a^*)^2 + (b^*)^2} \right] \text{ para } a^* > 0$$
$$360 - \left\{ \frac{180}{\pi} \times \cos \left[a / \frac{a}{\sqrt{(a^*)^2 + (b^*)^2}} \right] \right\} \text{ para } a^* < 0$$

O valor de croma reflete a saturação relativa, representada pela média de a^* e b^* . O croma foi calculado usando a seguinte equação:

$$Chroma = \sqrt{(a^*)^2 + (b^*)^2}$$

Essa análise detalhada dos valores de cor (L^* , a^* , b^* , Hue^* , $Chroma^*$) foi crucial para quantificar as características fenotípicas da cor das estruturas, permitindo uma avaliação objetiva de suas variações. A quantificação precisa desses parâmetros é essencial para estudos de herdabilidade e seleção de genótipos, pois facilita a identificação de características de cor que podem ser usadas como marcadores para o melhoramento genético.

2.8. Análise estatística e genética

A análise genética das coordenadas L^* , a^* , b^* , Hue^* e $Chroma^*$ foi conduzida com base em um modelo linear misto, conforme a seguinte estrutura:

$$Y_{ij} = \mu + B_j + G_i + \varepsilon_{ij}$$

Em que Y_{ij} representa o valor observado do genótipo i no bloco j ; μ é a constante geral do experimento; B_j é o efeito aleatório do bloco j ; G_i corresponde ao efeito do genótipo, decomposto em T_m (efeito fixo do tratamento comum) e T_{jm} (efeito aleatório do tratamento regular dentro do bloco j e o erro experimental ($\varepsilon_{ij} \sim NID(0, \sigma^2)$)).

A estimação dos componentes de variância foi realizada pelo método da máxima verossimilhança restrita (REML), enquanto as predições dos valores genéticos das linhagens foram obtidas por meio do estimador Best Linear Unbiased Prediction (BLUP). A significância dos efeitos fixos foi avaliada por meio do teste F, enquanto os efeitos aleatórios foram testados utilizando o Teste da Razão de Verossimilhança (LRT – *Likelihood Ratio Test*). A decomposição da soma de quadrados foi realizada pelo método do tipo III (ANOVA Tipo III), adequado para modelos desbalanceados e com interação entre fatores.

Os componentes de variância estimados incluíram a variância genética (σ_g^2) e a variância residual (σ_e^2). A herdabilidade no sentido amplo (H^2) foi estimada com base na seguinte fórmula:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

em que:

σ_g^2 é a variância genética,

σ_e^2 é a variância residual

A herdabilidade foi categorizada de acordo Kim *et al.* (2025) para três classes: 0,0-40% (Baixo), 60-79% (Médio) e mais de 80% (Alto).

Os valores de L*, a*, b*, Chroma* e Hue* obtidos para cada grupo de linhagens foram comparados estatisticamente. As diferenças entre as médias dos grupos foram testadas usando o teste t com o auxílio do pacote rstatix (Kassambara, 2023). Para a visualização dos resultados, foram construídos gráficos do tipo boxplot com o pacote ggpubr (Kassambara, 2023).

As análises foram realizadas no software R versão 4.4.1 (R Core Team, 2023). Para os modelos mistos e testes de significância, foram utilizados os pacotes lmerTest (Kuznetsova; Brockhoff; Christensen, 2017) e car (Fox; Weisberg, 2019).

3. RESULTADOS

3.1. Agrupamento das cores das estruturas

A avaliação visual das cores permitiu classificar as linhagens em três grupos para caule e pecíolo, quatro para frutos e folhas, e seis para sementes (Tabela 1). A análise de agrupamento hierárquico, baseada em processamento de imagens, foi utilizada apenas como representação das distâncias, sem aplicação de cortes para definir clusters. A coloração atribuída aos dendrogramas constitui um recurso visual de interpretação e não um resultado direto do algoritmo.

Tabela 1. Classificação visual direta das cores do galho, pecíolo, fruto, folha e semente (cor primária) de linhagens do Banco ativo de Germoplasma de Mamoneira.

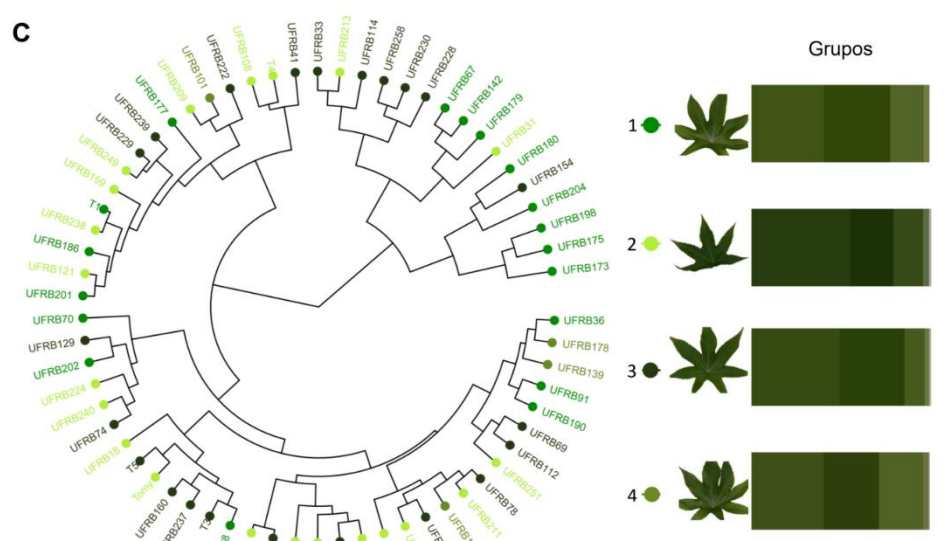
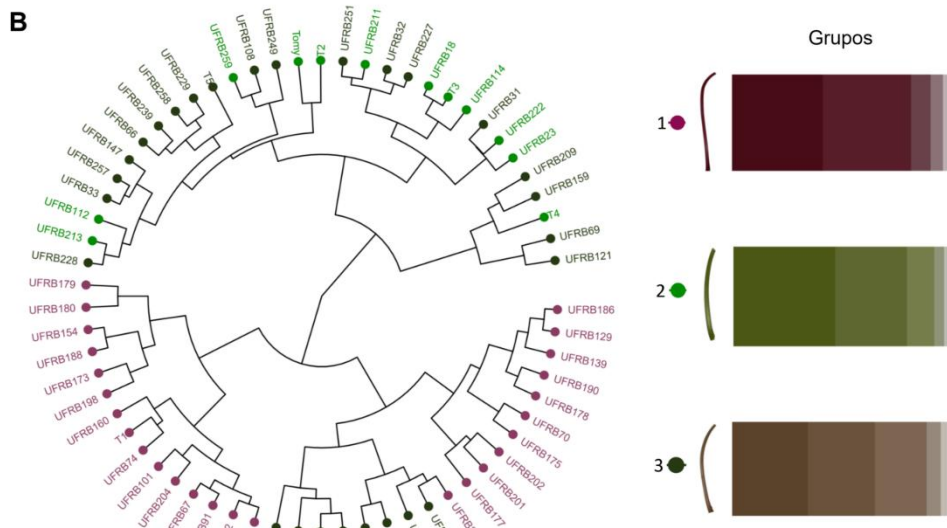
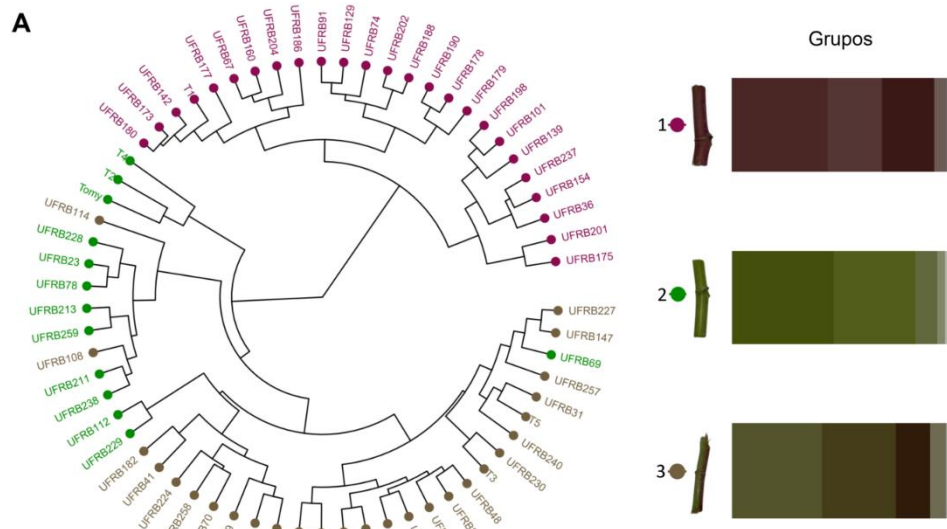
Grupos	Galho	Pecíolo	Fruto	Folha	Semente (cor primária)
1	Vermelho	Vermelho	Verde escura com acúleos roxos	Limbo Verde claro com nervura vermelha	Branca
2	Verde clara	Verde clara	Verde clara	Limbo Verde escuro	Marrom media
3	Verde- vermelho	Verde- vermelho	Verde escura	Limbo Verde claro com nervura verde clara	Marrom-escura
4			Verde clara com acúleos roxos	Limbo Verde escuro com nervura verde intermediária	Marrom avermelhada
5					Acinzentada
6					Preta

O dendrograma do galho (Figura 1A) evidenciou a organização hierárquica das linhagens segundo a similaridade cromática. As linhagens de galho vermelho agruparam-se em ramos próximos, enquanto aqueles de galho verde-claro

apresentaram distribuição mais dispersa. As linhagens com galho verde-avermelhado ocuparam posições intermediárias, sugerindo um gradiente contínuo. Do total de 65 linhagens, 38% apresentaram coloração vermelha, 20% verde-clara e 42% verde-avermelhada.

No pecíolo (Figura 1B), as linhagens vermelhas formaram um grupo coeso e distinto, em contraste com as de cor verde-clara e verde-avermelhada, que se distribuíram em ramos próximos, indicando maior similaridade entre si. Entre as 65 linhagens avaliados, 40% apresentaram pecíolos vermelhos, 18% verde-claros e 42% verde-avermelhados.

Figura 1. Dendrogramas de agrupamento hierárquico das 65 linhagens de mamoneira obtidos a partir da análise de imagens: (A) galho, (B) pecíolo e (C) folha.



Na folha (Figura 1C), não houve separação clara entre os quatro grupos de coloração identificados visualmente (verde-clara com nervura vermelha, verde-escura, verde-clara com nervura verde-clara e verde-escura com nervura verde-avermelhada). As linhagens desses grupos se distribuíram de forma intercalada, sugerindo variação cromática contínua. Entre as linhagens, 40% exibiram limbo verde-claro com nervura vermelha, 37% limbo verde-escuro com nervura verde-escura, 9% limbo verde-claro com nervura verde-clara e 14% limbo verde-escuro com nervura verde-avermelhada.

O dendrograma do fruto (Figura 2A) mostrou a formação de pequenos ramos coesos de acordo com a coloração (verde-escura com acúleos roxos, verde-clara, verde-escura e verde-clara com acúleos roxos), embora sem distinções amplas entre grupos. Do total de linhagens, 28% apresentaram frutos verde-escuros com acúleos roxos, 32% verde-claros, 34% verde-escuros e 6% verde-claros com acúleos roxos.

Na semente (Figura 2B), seis classes de coloração foram observadas: branca, marrom-clara, marrom-escura, marrom-avermelhada, acinzentada e preta. As classes preta e branca formaram grupos bem definidos, enquanto as marrons e acinzentadas exibiram distribuição mais complexa, com ocorrência de pequenos ramos por classe e misturas decorrentes de manchas secundárias. No total, 3% das linhagens apresentaram sementes brancas, 3% marrom-claras, 15% marrom-escuras, 6% marrom-avermelhadas, 8% acinzentadas e 65% pretas.

A

Grupos

1 2 3 4

B

Grupos

1 2 3 4 5 6

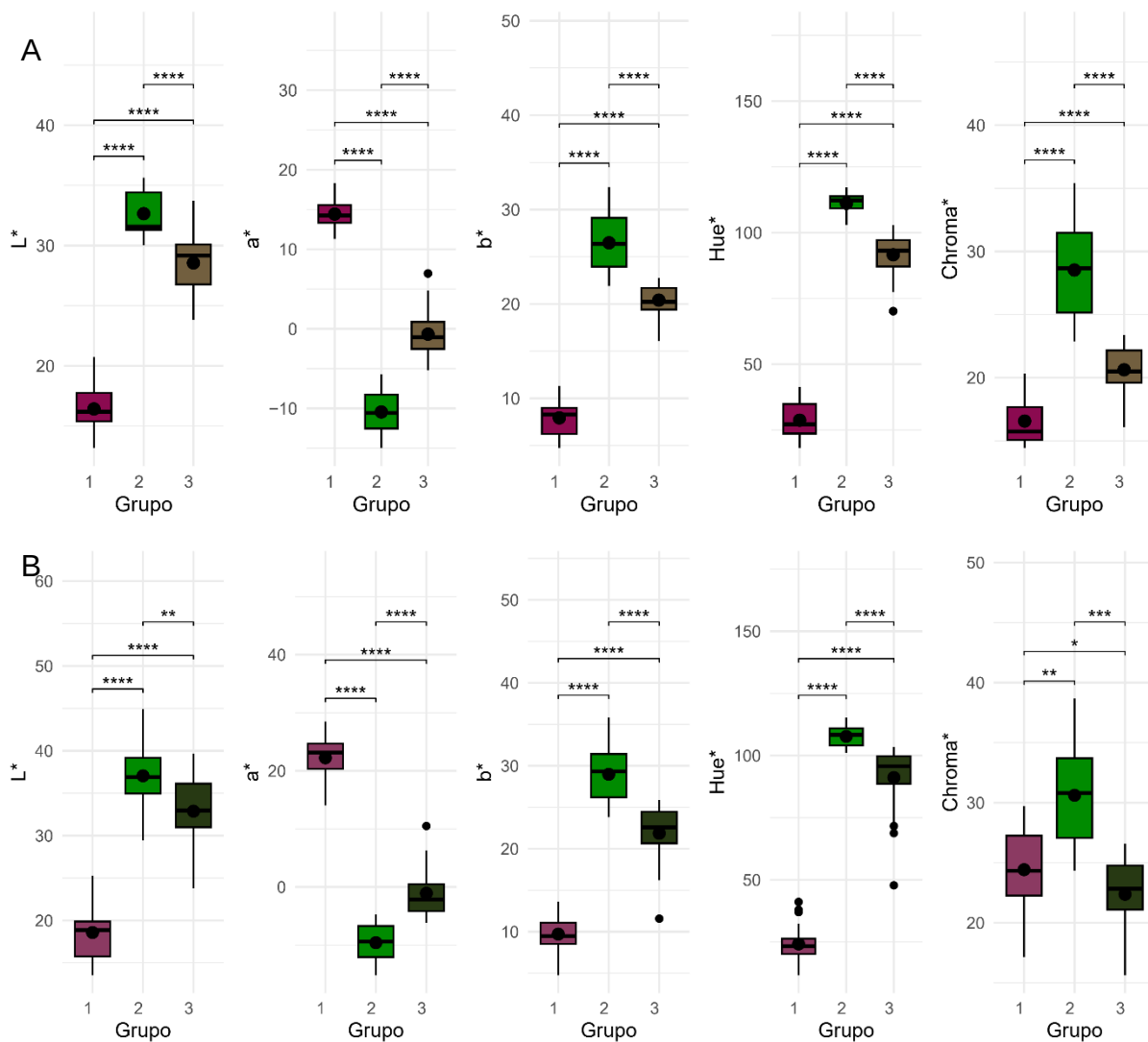
Os parâmetros colorimétricos (L^* , a^* , b^* , Hue* e Chroma*), obtidos a partir das imagens digitais das diferentes estruturas vegetais do BAG de mamoneira, evidenciaram padrões contrastantes de coloração entre galhos, folhas, frutos, pecíolos e sementes (Figuras 3 e 4). Diferenças significativas ($p < 0,05$ a $p < 0,0001$) foram detectadas entre as médias dos parâmetros em galhos, pecíolos, frutos e sementes. Para as folhas, entretanto, não foram observadas diferenças significativas em L^* e Hue*.

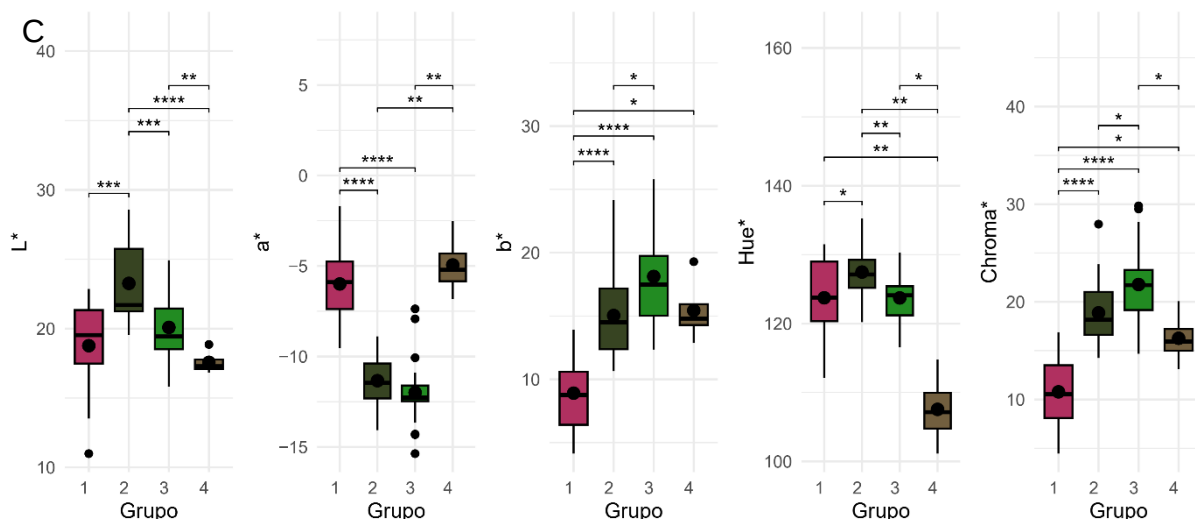
Nos galhos, três perfis foram identificados. O grupo 1 apresentou baixo valor de L^* (16,4) e a^* positivo elevado (14,4), caracterizando coloração escura de tonalidade vermelha intensa (Figura 3A). O grupo 2 exibiu o maior valor de L^* (32,7), a^* negativo (-10,5) e $Chroma^*$ elevado (28,5), correspondendo a galhos verde-intensos e saturados. O grupo 3 apresentou valores de a^* próximos de zero (-0,66) e Hue^* elevado (91,6), consistentes com coloração intermediária verde-avermelhada.

Para os pecíolos, três padrões cromáticos foram observados (Figura 3B). O grupo 1 apresentou a^* positivo elevado (22,2) e Hue^* reduzido (24,2), caracterizando pecíolos vermelho-intensos. O grupo 2 exibiu a^* próximo de zero (-9,6) e Hue^* = 108, consistentes com pecíolos verde-vivos. O grupo 3 apresentou valores de a^* próximos de zero (-1,08) e Hue^* = 91,1, indicando tonalidade verde-avermelhada.

Nas folhas, observou-se uniformidade cromática, com valores de a^* fortemente negativos (-17,4 a -20,2) e Hue^* estáveis (123–125), confirmando pigmentação predominantemente verde (Figura 3C). Os valores de $Chroma^*$ elevados (32,3–36,2) indicaram alta vivacidade cromática, reforçando a homogeneidade da coloração.

Figura 3. Diferença entre os grupos de linhagens de mamoneira com base em parâmetros colorimétricos cor (L^* , a^* , b^* , Hue* e Chroma*) para (A) Galho, (B) Pecíolo e (C) Folha.



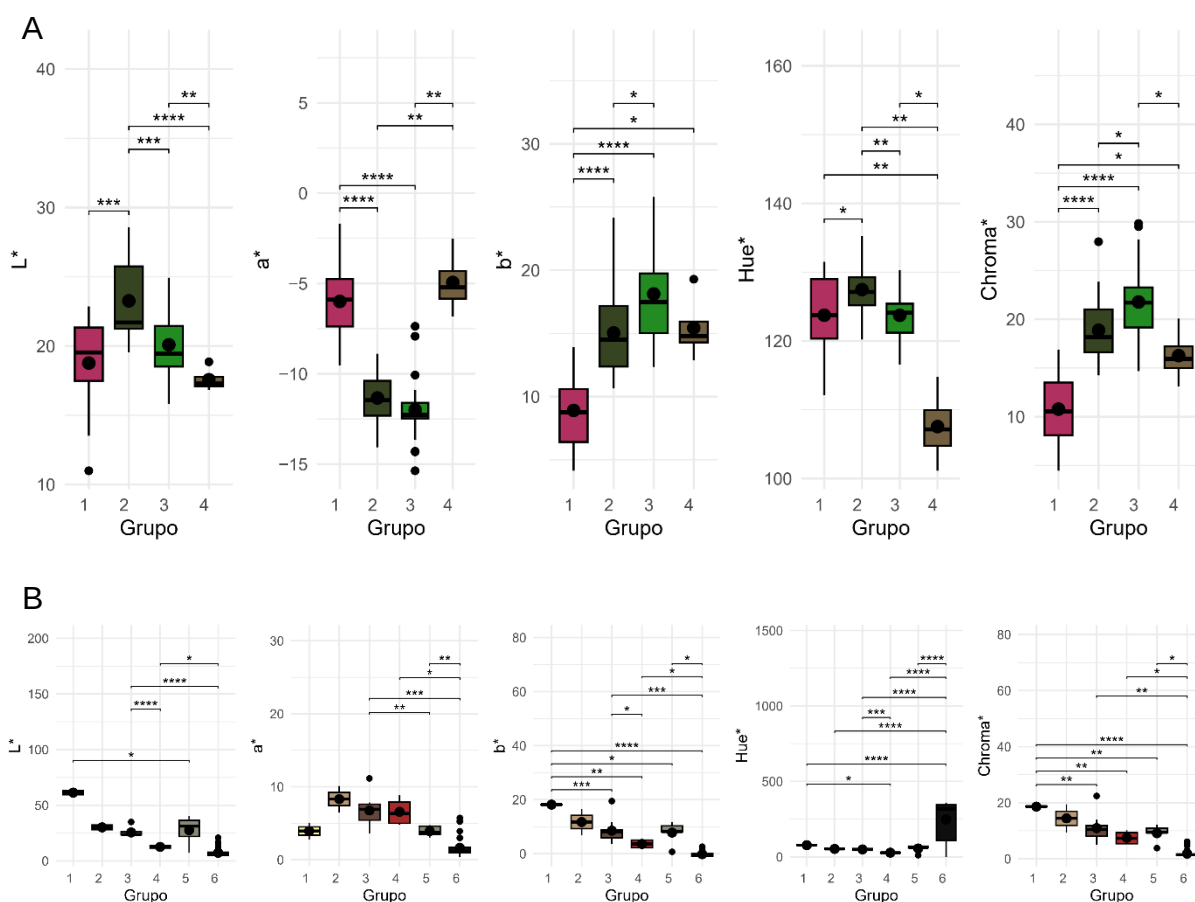


Teste t realizado para definir as diferenças mínimas significativas entre os pares dos grupos. ****, ***, **, *, significativo a 0,01, 0,1, 1 e 5% de probabilidade.

Nos frutos, os valores de L* variaram entre 17,6 e 23,3, indicando tonalidade escura (Figura 4A). Os valores negativos de a* (-4,9 a -12,0) e Hue* entre 108 e 127 corresponderam a diferentes tons de verde. A amplitude de variação do Chroma* (10,8–21,8) refletiu diferenças de saturação, provavelmente associadas à coloração dos acúleos.

As sementes apresentaram a maior diversidade de coloração entre as estruturas (Figura 4B). O grupo 1 apresentou L* elevado (61,3) e Hue* = 77,8, caracterizando sementes claras e amareladas. Os grupos 2 e 3 exibiram L* intermediário (25,6–30,2), a* positivo (6,8–8,3) e Hue* entre 49 e 52, correspondendo a sementes marrom-avermelhadas. O grupo 4 apresentou L* baixo (12,7) e Hue* reduzido (27,8), caracterizando sementes vermelho-escuras e opacas. O grupo 5 exibiu L* intermediário (27,6) e Hue* = 55,7, correspondendo a sementes acinzentadas. Por fim, o grupo 6 apresentou L* muito baixo (7,3), Hue* elevado (248) e Chroma* reduzido (1,9), indicando sementes pretas de baixa saturação.

Figura 4. Diferença entre os grupos de linhagens de mamoneira com base em parâmetros colorimétricos cor (L*, a*, b*, Hue* e Chroma*) para (A) Fruto, (B) Sementes.



Teste t realizado para definir as diferenças mínimas significativas entre os pares dos grupos. ****, ***, **, *, significativo a 0,01, 0,1, 1 e 5% de probabilidade.

3.3. Herdabilidade dos atributos de cor

A análise de deviance indicou variâncias genéticas significativas entre as linhagens para os componentes de cor (L^* , a^* , b^* , Hue^* e $Chroma^*$) em todas as estruturas avaliadas, com exceção das folhas. Em contrapartida, os efeitos de posição (testemunhas e blocos) apresentaram impacto reduzido (Tabela 2).

As estimativas de herdabilidade no sentido amplo (H^2) variaram conforme a estrutura analisada. Nos galhos, observaram-se valores elevados para L^* (0,93), a^* (0,92), b^* (0,91) e Hue^* (0,97), e médios para $Chroma^*$ (0,75). No pecíolo, a herdabilidade também foi elevada para L^* (0,89), a^* (0,96), b^* (0,89) e Hue^* (0,97), e média para $Chroma^*$ (0,67). Nos frutos, a herdabilidade foi média para a^* (0,73) e $Chroma^*$ (0,61), e baixa para L^* (0,49), b^* (0,55) e Hue^* (0,16). Para as sementes, os valores foram elevados para a^* (0,87) e Hue^* (1,00), e médios para L^* (0,67), b^* (0,61) e $Chroma^*$ (0,60). Em contraste, nas folhas, todas as variáveis de cor

apresentaram herdabilidade baixa ($L^* = 0,23$; $a^* = 0,39$; $b^* = 0,56$; $Hue^* = 0,28$; $Chroma^* = 0,56$).

Tabela 2. Componentes quadráticos e herdabilidade para os parâmetros de cor (L^* , a^* , b^* , Hue^* e $Chroma^*$) em galho, pecíolo, fruto, folha e sementes de linhagens do Banco ativo de Germoplasma de Mamoneira.

Estrutura	Efeito e herdabilidade	L^*	a^*	b^*	Hue^*	$Chroma^*$
Galho	Testemunha ^f	3,22	4,62*	6,42*	2,40	14,20***
	Bloco ^f	0,73	1,45	0,44	2,12	0,18
	σ_g^2	46,64***	87,80***	50,51***	1195,00***	15,83***
	σ^2	3,44	7,77	4,9	42,41	5,38
	H^2	0,93	0,92	0,91	0,97	0,75
Pecíolo	Testemunha ^f	3,23	5,02*	7,46**	3,19	11,84**
	Bloco ^f	2,14	0,38	0,33	0,73	0,15
	σ_g^2	65,37***	169,67***	51,34***	1312,91***	12,69***
	σ^2	7,77	7,48	6,65	34,47	6,37
	H^2	0,89	0,96	0,89	0,97	0,67
Fruto	Testemunha ^f	6,90*	2,36	2,65	0,06	2,77
	Bloco ^f	3*	1,06	1,7	0,47	1,68
	σ_g^2	6,22***	7,35***	11,17**	6,22	16,57***
	σ^2	6,41	2,75	9,23	33,29	10,55
	H^2	0,49	0,73	0,55	0,16	0,61
Folha	Testemunha ^f	0,27	4,14	2,59	0,26	3,03
	Bloco ^f	0,72	0,14	2,27	4,92**	1,37
	σ_g^2	5,21	1,68	9,26	1,54	10,53
	σ^2	17,26	2,58	7,17	3,87	8,22
	H^2	0,23	0,39	0,56	0,28	0,56
Semente	Testemunha ^f	5,53***	4,15**	9,54***	0,71	9,22***
	Bloco ^f	1,22	1,04	2,39	0,64	2,52
	σ_g^2	111,77***	3,81*	11,07*	22719,96***	10,98*
	σ^2	16,97	1,84	7,18	75,37	7,33
	H^2	0,87	0,67	0,61	99,67	0,60

σ_g^2 : Variância genética de linhagens; σ^2 : Variância residual; h^2 : herdabilidade no sentido amplo; ^f efeito fixo; ^a efeito aleatório. Teste F para efeitos fixos e Teste da Razão de Verossimilhança (LRT) para efeitos aleatórios. ****, ***, **, *, significativo a 0,01, 0,1, 1 e 5% de probabilidade.

4. DISCUSSÃO

A caracterização quantitativa da cor por visão computacional demonstrou-se eficaz como ferramenta complementar às avaliações visuais, reduzindo a subjetividade inerente às observações diretas.

A análise do perfil quantitativo no espaço RGB, baseada na distância EMD, revelou agrupamentos hierárquicos consistentes entre as estruturas vegetais, com contrastes claros entre linhagens de coloração vermelha, verde e verde-

avermelhada em galhos e pecíolos. Nos frutos e sementes, a presença de cores secundárias também contribuiu para a formação dos agrupamentos. Resultados análogos, nos quais histogramas de cor extraídos de pixels combinados à métrica EMD permitiram capturar variações cromáticas relevantes, foram reportados em leguminosas como flores e sementes de feijão (*Phaseolus vulgaris*) e amendoim forrageiro (*Arachis pintoï*) (Rodríguez *et al.*, 2024), bem como em flores de orquídeas (*Ophrys aveyronensis*) (Gibert *et al.*, 2025). De forma semelhante, Isaza *et al.* (2018) observaram que diferenças em valores RGB, convertidos em histogramas de cor de sementes de plantas silvestres de mamoneira provenientes de dez localidades geográficas distintas, foram suficientes para discriminar grupos, corroborando os achados do presente estudo em linhagens do BAG de mamoneira.

Tradicionalmente, a coloração é utilizada como descritor morfológico em estudos de caracterização de germoplasma de mamoneira (Anjani, 2012; Silva *et al.*, 2019; Wahibah *et al.*, 2020), mas a avaliação visual depende da experiência do observador e pode resultar em inconsistências entre coleções. Nesse contexto, a fenotipagem digital tem ampliado o valor desses descritores clássicos, ao permitir quantificação objetiva, em larga escala e reproduzível das cores (Halcro *et al.*, 2020; Underhill *et al.*, 2020; Hemingson *et al.*, 2024).

A avaliação no espaço CIELAB possibilitou quantificar objetivamente os parâmetros colorimétricos (L^* , a^* , b^* , Hue^* e $Chroma^*$), evidenciando diferenças significativas em galhos, pecíolos, frutos e sementes. A ausência de diferenças para L^* e Hue^* nas folhas, associada aos baixos valores de herdabilidade obtidos, indica maior influência ambiental na pigmentação foliar. Isso indica que a coloração das folhas pode ser um caráter pouco informativo para diferenciar linhagens.

Sob a perspectiva ecológica, a variação de cor observada nas diferentes estruturas pode estar associada a funções adaptativas. Pigmentos antociânicos em folhas, galhos e pecíolos estão relacionados à fotoproteção contra radiação UV e à defesa contra herbivoria (Mahdavian *et al.*, 2024; Cornelissen *et al.*, 2025). A diversidade de padrões cromáticos em sementes pode refletir estratégias de mimetismo (Myczko *et al.*, 2015), enquanto a uniformidade de coloração nas folhas sugere maior restrição funcional, possivelmente ligada à eficiência fotossintética (Su *et al.*, 2023; Li *et al.*, 2023).

As estimativas de herdabilidade variaram de baixas a altas conforme a estrutura analisada, com destaque para valores consistentemente elevados em

galhos, pecíolos e sementes. Nas folhas, os valores baixos observados em mamoneira (0,23–0,56) contrastam com herdabilidades elevadas reportadas em feijão ($L^* = 0,88$; Chroma = 0,87; Celebioglu *et al.*, 2024) e intermediárias em *Coleus scutellarioides* (Li *et al.*, 2021), reforçando que a pigmentação foliar em mamoneira apresenta baixa herdabilidade, indicando que a maior parte da variação observada se deve a fatores ambientais. Para os frutos, herdabilidades médias a baixas (0,16–0,73) foram consistentes com os intervalos relatados em espécies perenes e anuais, como pêssogo (*Prunus persica*) (0,60–0,77; Dini *et al.*, 2019), manga (*Mangifera indica*) (0,28–0,52; Costa *et al.*, 2023) e abóbora (*Cucurbita* spp.) (0,12–0,49; Itle *et al.*, 2022). Esses resultados indicam que a variação observada na coloração do fruto é explicada em maior parte por fatores não genéticos. Já nas sementes, os valores elevados obtidos (0,60–1,00), com destaque para Hue* (99,67), estão em consonância com herdabilidades próximas da unidade descritas em feijão (0,98–0,99; Sadohara *et al.*, 2021) e linho (*Linum Usitatissimum*) (0,88–0,91; Abtahi *et al.*, 2022), evidenciando forte controle genético e a utilidade da coloração da semente como descritor estável em programas de melhoramento e caracterização de germoplasma de mamoneira.

De forma geral, a combinação de avaliação visual, agrupamento hierárquico no espaço RGB e quantificação no espaço CIELAB mostrou-se eficiente para caracterizar a diversidade cromática de linhagens de mamoneira. A integração dessas abordagens não apenas amplia o entendimento do controle genético da pigmentação, mas também fornece descritores fenotípicos de alta herdabilidade, particularmente em sementes, pecíolos e galhos. Esses resultados indicam que a fenotipagem digital da cor pode ser incorporada a estratégias de caracterização e seleção assistida, fortalecendo o uso do BAG de mamoneira tanto em conservação quanto em programas de melhoramento.

Entretanto, alguns aspectos devem ser considerados. Primeiramente, não foram definidos grupos formais nos dendrogramas, sendo a coloração visual utilizada como recurso interpretativo; futuros trabalhos poderão integrar métodos estatísticos de determinação de clusters para complementar a análise hierárquica. Além disso, as análises foram conduzidas por estrutura de forma independente; abordagens multivariadas poderiam explorar de forma mais integrada a associação entre coloração de diferentes órgãos. Por fim, a variabilidade observada em folhas e frutos sugere maior influência ambiental sobre esses caracteres, reforçando a

importância de estudos em ambientes controlados para confirmar sua estabilidade.

5. CONCLUSÃO

A caracterização quantitativa da cor utilizando visão computacional demonstrou-se uma abordagem eficaz e complementar às avaliações visuais, reduzindo a subjetividade das observações diretas. A combinação de análise de agrupamentos hierárquicos no espaço RGB, baseada na métrica EMD, e de mensurações objetivas no espaço CIELAB permitiu identificar contrastes cromáticos consistentes entre linhagens e estruturas vegetais, incluindo galhos, pecíolos, frutos e sementes.

As elevadas estimativas de herdabilidade e a variância genética significativa observadas para a maioria das variáveis colorimétricas evidenciam forte controle genético sobre esses caracteres, indicando potencial de utilização direta em programas de melhoramento da mamoneira. Nesse contexto, as variáveis de cor se configuram como marcadores fenotípicos robustos, capazes de auxiliar a seleção de genótipos e contribuir para a identificação e manutenção da diversidade cromática em bancos de germoplasma.

6. REFERÊNCIAS

ABTAHI, M.; MIRLONI, A.; SHARIF-MOGHADDAM, N.; ATAIL, E. Revealing seed color variation and their possible association with yield and quality traits in a diversity panel of flax (*Linum Usitatissimum* L.). **Frontiers in Plant Science**, v. 13, p. 1038079, 2022. DOI: <https://doi.org/10.3389/fpls.2022.1038079>

ANJANI, K. J. I. C. Castor genetic resources: A primary gene pool for exploitation. **Industrial Crops and Products**, v. 35, n. 1, p. 1-14, 2012. DOI: <https://doi.org/10.1016/j.indcrop.2011.06.011>

BAJGAIN, P.; LI, C.; ANDERSON, J. A. Genome-wide association mapping and genomic prediction for kernel color traits in intermediate wheatgrass (*Thinopyrum intermedium*). **BMC plant biology**, v. 22, n. 1, p. 218, 2022. DOI: <https://doi.org/10.1186/s12870-022-03616-7>

BRASIL. Instruções para execução dos ensaios de distinguibilidade, homogeneidade e estabilidade de cultivares de mamona (*Ricinus communis* L.). 2008. DOU nº 147, de 01/08/2008, seção 1, p. 14-15.

BOSTEN, Jenny M. Do you see what I see? Diversity in human color perception. **Annual review of vision science**, v. 8, n. 1, p. 101-133, 2022. DOI: <https://doi.org/10.1146/annurev-vision-093020-112820>

CELEBIOGLU, B.; MYERS, J. R.; HART, J. P.; PORCH, T.; GRIFFITHS, P. Phenotypic variability for leaf and pod color within the Snap bean association panel. **Journal of the American Society for Horticultural Science**, v. 149, n. 1, p. 15-26, 2024. DOI: 10.21273/JASHS05326-23

CORNELISSEN, T.; SILVEIRA, F. A.; GOMES, S. V.; LOPEZ-GOLDAR, X.; MARTIN-EBERHARDT, S.; WETZEL, W. Signaling defenses with color: a meta-analysis of leaf color variation, palatability, and herbivore damage. **New Phytologist**, 2025. DOI: <https://doi.org/10.1111/nph.70243>

COSTA, C. D. S. R.; COSTA, A. E. D. S.; NETO, F. P. L.; LIMA, M. A. C.; MARTINS, L. S. S.; MUSSER, R. D. S. Repeatability coefficient for fruit quality and selection of mango hybrids using REML/BLUP analysis. **Euphytica**, v. 219, n. 11, p. 120, 2023. DOI: <https://doi.org/10.1007/s10681-023-03249-3>

DINI, M.; RASEIRA, M. C. B.; SCARIOTTO, S.; CARRA, B.; ABREU, E. S.; MELLO-FARIAS, P.; CANTILLANO, R. F. F. Color Shade Heritability of Peach Flesh. **Journal of Agricultural Science**, v. 11, n. 8, p. 236-247, 2019. DOI: <https://doi.org/10.5539/jas.v11n8p236>

DU, J.; FAN, J.; WANG, C.; LU, X.; ZHANG, Y.; WEN, W.; LIAO, S.; YANG, X.; GUO, X.; ZHAO, C. Greenhouse-based vegetable high-throughput phenotyping platform and trait evaluation for large-scale lettuces. **Computers and Electronics in Agriculture**, v. 186, p. 106193, 2021. DOI: <https://doi.org/10.1016/j.compag.2021.106193>

FOX, J.; WEISBERG, S. **An R Companion to Applied Regression**, Third edition.

Sage, Thousand Oaks CA. <https://www.john-fox.ca/Companion/>. 2018.

GIBERT, A.; SCHATZ, B.; BUSCAIL, R.; NGUYEN, D.; BAGUETTE, M.; BARTHES, N.; BERTRAND, J. A. Floral phenotypic divergence and genomic insights in an *Ophrys* orchid: unraveling early speciation processes. **New Phytologist**, v. 245, n. 2, p. 849-868, 2025. DOI: <https://doi.org/10.1111/nph.20190>

GU, Z; GU, L; EILS, R; SCHLESNER, M.; BRORS, B. "circlize implements and enhances circular visualization in R." **Bioinformatics**, v.30, p. 2811-2812, 2014. DOI: <https://doi.org/10.1093/bioinformatics/btu393>

HALCRO, K.; MCNABB, K.; LOCKINGER, A.; SOCQUET-JUGLARD, D.; BETT, K. E.; NOBLE, S. D. The BELT and phenoSEED platforms: shape and colour phenotyping of seed samples. **Plant Methods**, v. 16, n. 1, p. 49, 2020. DOI: <https://doi.org/10.1186/s13007-020-00591-8>

HEMINGSON, C. R.; COWMAN, P. F.; BELLWOOD, D. R. Analysing biological colour patterns from digital images: An introduction to the current toolbox. **Ecology and Evolution**, v. 14, n. 3, p. e11045, 2024. DOI: <https://doi.org/10.1002/ece3.11045>

ISAZA, C.; ANAYA, K.; PAZ, J. Z.; VASCO-LEAL, J. F.; HERNANDEZ-RIOS, I.; MOSQUERA-ARTAMONOV, J. D. Image analysis and data mining techniques for classification of morphological and color features for seeds of the wild castor oil plant (*Ricinus communis* L.). **Multimedia Tools and Applications**, v. 77, n. 2, p. 2593-2610, 2018. DOI: <https://doi.org/10.1007/s11042-017-4438-y>

ITLÉ, R. A.; KABELKA, E. A.; OLMSTEAD, J. W. Heritability estimates of L* a* b* color space values in winter squash (*Cucurbita* spp.). **HortScience**, v. 57, n. 2, p. 202-214, 2022. DOI: [10.21273/HORTSCI15961-21](https://doi.org/10.21273/HORTSCI15961-21)

KARIMIPOUR, H.; WITZEL, C. Colour expectations across illumination changes. **Vision Research**, v. 222, p. 108451, 2024. DOI: <https://doi.org/10.1016/j.visres.2024.108451>

KASSAMBARA, A. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.1, 2025. Disponível em: <https://rpkgs.datanovia.com/ggpubr/>. Acesso em: 21 de agosto de 2025.

KIM, J. M.; LEE, J. W.; KIM, D. J.; LYU, J. I.; BAEK, J.; HA, B. K.; KWON, S. J. Image-based GWAS identifies the genetic architecture of seed-related traits in a soybean mutant population. **Molecular Breeding**, v. 45, n. 8, p. 67, 2025. DOI: <https://doi.org/10.1007/s11032-025-01584-y>

KUZNETSOVA, A.; BROCKHOFF, P. B.; CHRISTENSEN, R. H. B. lmerTest package: tests in linear mixed effects models. **Journal of statistical software**, v. 82, p. 1-26, 2017. DOI: [10.18637/jss.v082.i13](https://doi.org/10.18637/jss.v082.i13)

LI, G.; LIANG, H.; REN, X.; MA, W.; LU, Y.; ZHANG, Z.; WANG, Z.; ZHAO, T.; ZHAO, J. Transcriptome Analysis Reveals Association of Photosynthesis and Phytohormone Pathways with Leaf Color in Chinese Cabbage. **Agronomy**, v. 13, n.

9, p. 2273, 2023. DOI: <https://doi.org/10.3390/agronomy13092273>

LI, J.; CHEN, J.; ZHANG, Q.; YU, P.; ZHOU, Y.; JIA, G. The composition of anthocyanins and carotenoids influenced the flower color heredity in Asiatic hybrid lilies. **Horticulturae**, v. 8, n. 12, p. 1206, 2022. DOI: <https://doi.org/10.3390/horticulturae8121206>

LI, M.; CONEVA, V.; ROBBINS, K. R.; CLARK, D.; CHITWOOD, D.; FRANK, M. Quantitative dissection of color patterning in the foliar ornamental coleus. **Plant Physiology**, v. 187, n. 3, p. 1310-1324, 2021. DOI: <https://doi.org/10.1093/plphys/kiab393>

LOARCA, J.; WIESNER-HANKS, T.; LOPEZ-MORENO, H.; MAULE, A. F.; LIOU, M.; TORRES-MERAZ, M. A.; DIAZ-GARCIA, L.; JHONSON-CICALESE, J.; NEYHART, J.; POLASHOCK, J.; SIDELE, G. M.; STROCK, C. F.; BEIL, C.T.; SHEEHAN, M. J.; LORIZZO, M.; ATUCHA, A.; ZALAPA, J. BerryPortraits: Phenotyping Of Ripening Traits cranberry (*Vaccinium macrocarpon* Ait.) with YOLOv8. **Plant Methods**, v. 20, n. 1, p. 172, 2024. DOI: <https://doi.org/10.1186/s13007-024-01285-1>

LU, C.; LI, Y.; WANG, J.; QU, J.; CHEN, Y.; CHEN, X.; HUANG, H.; DAI, S. Flower color classification and correlation between color space values with pigments in potted multiflora chrysanthemum. **Scientia Horticulturae**, v. 283, p. 110082, 2021. DOI: <https://doi.org/10.1016/j.scienta.2021.110082>

MAHDAVIAN, K. Effects of ultraviolet radiation on plants and their protective mechanisms. **Russian Journal of Plant Physiology**, v. 71, n. 6, p. 184, 2024. DOI: <https://doi.org/10.1134/S1021443724607481>

MATHIAZHAGAN, M.; CHINNAIYAN, V.; RAO, D. V. S.; SHIVASHANKARA, K. S.; RAVISHANKAR, K. V. Estimation of Pulp Pigment Content Using Color Space Values in Guava. **Applied Fruit Science**, v. 66, n. 5, p. 1997-2005, 2024. DOI: <https://doi.org/10.1007/s10341-024-01157-x>

MYCZKO, Ł.; SKÓRKA, P.; DYLEWSKI, Ł.; SPARKS, T. H.; TRYJANOWSKI, P. Color mimicry of empty seeds influences the probability of predation by birds. **Ecosphere**, v. 6, n. 10, p. 1-7, 2015. DOI: <https://doi.org/10.1890/ES15-00055.1>

NOUR, I. H.; ALHADEAD, K.; ELLMOUNI, F. Y.; BADR, R.; SAAD, T. I.; EL-BANHAWY, A.; ABDEL RAHMAN, S. M. Morphological, anatomical and chemical characterization of *Ricinus communis* L. (Euphorbiaceae). **Agronomy**, v. 13, n. 4, p. 985, 2023. DOI: <https://doi.org/10.3390/agronomy13040985>

PARADIS, E.; SCHLIEP, K. "ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R." **Bioinformatics**, v.35, p. 526-528, 2019. DOI: <https://doi.org/10.1093/bioinformatics/bty633>

R Core Team. R: A language and environment for statistical computing. **Version 4.4.1, R Foundation for Statistical Computing, Vienna, Austria**, 2023. Disponível em: <https://www.R-project.org/>. Acesso em: 22 de agosto de 2025.

REYNOLDS, D.; SINGH, V. Characterization of human lightness discrimination thresholds for independent spectral variations. **i-Perception**, v. 15, n. 5, p. 20416695241274662, 2024. DOI: <https://doi.org/10.1177/20416695241274662>

RIBEIRO, L. P.; SANTOS, D. M. B.; NETO, L. A. L.; NETO, L. R. S.; BARBOSA, M. F.; CUNHA, T. J. F. Levantamento detalhado dos solos, capacidade de uso, e classificação de terras para irrigação da estação de plasticultura da Universidade Federal da Bahia/Politeo em Cruz das Almas (BA). **Revista Brasileira de Ciência do Solo, Campinas**, v. 19, n. 1, p. 105-113, 1995.

RODRÍGUEZ, D. F. C.; GUZMAN, J. J. G.; GIL, J. G. R.; WENZL, P.; URBAN, M. O. Digital descriptors sharpen classical descriptors, for improving genebank accession management: A case study on *Arachis* spp. and *Phaseolus* spp. **Plos one**, v.19, n.5, p. e0302158, 2024. DOI: <https://doi.org/10.1371/journal.pone.0302158>

SADOHARA, R.; LONG, Y.; IZQUIERDO, P.; URREA, C. A.; MORRIS, D.; CICHY, K. Seed coat color genetics and genotype× environment effects in yellow beans via machine-learning and genome-wide association. **The Plant Genome**, v. 15, n. 1, p. e20173, 2021. DOI: <https://doi.org/10.1002/tpg2.20173>

SILVA, A. R. D.; SILVA, S. A.; SANTOS, L. A.; SOUZA, D. R.; ARAUJO, G. M.; DANTAS, J. L. L.; LEITE, E. S.; DANTAS, A. C. V. L. Characterization and performance of castor bean lineages and parents at the UFRB germplasm bank. **PloS one**, v. 14, n. 1, p. e0209335, 2019. DOI: <https://doi.org/10.1371/journal.pone.0209335>

SU, X.; YUE, X.; KONG, M.; XIE, Z.; YAN, J.; MA, W.; WANG, Y.; ZHAO, J.; ZHANG, X.; LIU, M. Leaf color classification and expression analysis of photosynthesis-related genes in inbred lines of Chinese cabbage displaying minor variations in dark-green leaves. **Plants**, v. 12, n. 11, p. 2124, 2023. DOI: <https://doi.org/10.3390/plants12112124>

UNDERHILL, A. N.; HIRSCH, C. D.; CLARK, M. D. Evaluating and mapping grape color using image-based phenotyping. **Plant Phenomics**, 2020. DOI: 10.34133/2020/8086309

URBANEK, S. jpeg: Read and write JPEG images. Versão 0.1-11. RForge, 2025. Disponível em: <https://www.rforge.net/jpeg/>. Acesso em: 27 de setembro de 2025.

URBANEK, S. png: Read and write PNG images. Versão 0.1-8. CRAN, 2022. DOI: 10.32614/CRAN.package.png.

WAHIBAH, N. N.; YAHYA, V. J.; PERDANA, M. A.; BUDIONO, R. Morphological variation of castor bean (*Ricinus communis* L.) on peatland area in Kepulauan Meranti Riau Indonesia. In: **Journal of Physics: Conference Series**. IOP Publishing, p. 012028, 2020. DOI: 10.1088/1742-6596/1655/1/012028

WAN, W.; JIA, F.; LIU, Z.; SUN, W.; ZHANG, X.; SU, J.; GUAN, Z.; CHEN, F.; ZHANG, F.; FANG, W. Quantitative evaluation and genome-wide association studies of chrysanthemum flower color. **Scientia Horticulturae**, v. 338, p. 113561, 2024. DOI: <https://doi.org/10.1016/j.scienta.2024.113561>

WELLER, H. I.; WESTNEAT, M. W. Quantitative color profiling of digital images with earth mover's distance using the R package colordistance. **PeerJ**, v. 7, p. e6398, 2019. DOI: <https://doi.org/10.7717/peerj.6398>

WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K.; VAUGHAN, D. dplyr: A Grammar of Data Manipulation. Versão 1.1.4. CRAN, 2023. DOI: 10.32614/CRAN.package.dplyr.

YU, G. Using ggtree to visualize data on tree-like structures. **Current Protocols in Bioinformatics**, p.69:e96, 2020. DOI: <https://doi.org/10.1002/cpbi.96>

ZEILEIS, A.; FISHER, J. C.; HORNIK, K.; IHAKA, R.; MCWHITE, C. D.; MURRELL, P.; STAUFFER, R.; WILKE, C. O. "colorspace: A Toolbox for Manipulating and Assessing Colors and Palettes." **Journal of Statistical Software**, v. 96, n. 1, p. 1–49, 2020. Doi:10.18637/jss.v096.i01.