

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL  
CÂMPUS DE CHAPADÃO DO SUL  
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

IZADORA ARAUJO OLIVEIRA

**RELAÇÃO ENTRE DADOS HIPERESPECTRAIS E A COMPOSIÇÃO DE  
AMINOÁCIDOS EM GENÓTIPOS DE SOJA**

CHAPADÃO DO SUL – MS

2025

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL  
CÂMPUS DE CHAPADÃO DO SUL  
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

IZADORA ARAUJO OLIVEIRA

**RELAÇÃO ENTRE DADOS HIPERESPECTRAIS E A COMPOSIÇÃO DE  
AMINOÁCIDOS EM GENÓTIPOS DE SOJA**

Orientador(a): Prof (a). Dr(a) Ana Carina da Silva Cândido Seron

Co-Orientador: Prof (a). Dr(a) Elisângela de Souza Loureiro

Dissertação apresentada à Universidade Federal de Mato Grosso do Sul, como parte dos requisitos para obtenção do título de Mestre em Agronomia, área de concentração: Produção Vegetal.

CHAPADÃO DO SUL – MS

2025



PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

CERTIFICADO DE APROVAÇÃO

**DISCENTE:** Izadora Araujo Oliveira

**ORIENTADORA:** Dra. Ana Carina da Silva Candido Seron

**TÍTULO:** Relação entre Dados Hiperspectrais e a Composição de Aminoácidos em Genótipos de Soja.

**AVALIADORES:**

Prof. Dra. Ana Carina da Silva Candido Seron

Prof. Dra. Dthenifer Cordeiro Santan

Prof. Dra. Mayara Fávero Cotrim

Chapadão do Sul, 16 de outubro de 2025.

NOTA  
MÁXIMA  
NO MEC

UFMS  
É 10!!!



Documento assinado eletronicamente por **Ana Carina da Silva Candido, Professora do Magistério Superior**, em 16/10/2025, às 14:46, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA  
MÁXIMA  
NO MEC

UFMS  
É 10!!!



Documento assinado eletronicamente por **Mayara Fávero Cotrim, Usuário Externo**, em 16/10/2025, às 15:04, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA  
MÁXIMA  
NO MEC

UFMS  
É 10!!!



Documento assinado eletronicamente por **Dthenifer Cordeiro Santana, Usuário Externo**, em 16/10/2025, às 15:09, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufms.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufms.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **5960388** e o código CRC **D4A2D828**.

---

## COLEGIADO DE PÓS-GRADUAÇÃO EM AGRONOMIA

Avenida Engenheiro Douglas Ribeiro Pantaleão, nº 5167 Fone:

(67)3562-6351

CEP 79560-000 - Chapadão do Sul - MS

---

**Referência:** Processo nº 23455.000025/2025-27  
5960388

SEI nº

## **DEDICATÓRIA**

Primeiramente à Deus, por me capacitar e sempre guiar meu caminho com fé e perseverança.  
Aos meus pais Zilda Eleusa Araujo oliveira e Adalberto Costa de Oliveira, minha Irmã Doraiza Araujo Oliveira e meu esposo Eduardo Pedroso Ferreira, eles são absolutamente tudo para mim.

## **AGRADECIMENTOS**

À minha Mãe Zilda Eleusa Araujo oliveira e meu Pai Adalberto Costa de Oliveira, minha Irmã Doraiza Araujo Oliveira pelo esforço, compreensão e apoio em todas as fases da minha vida. Mesmo morando em outro estado, são presentes em minha vida sempre.

Ao meu esposo Eduardo Pedroso Ferreira, por sempre estar ao meu lado me apoiando e incentivando a alcançar meus objetivos.

Minhas colegas de trabalho e estudo Marina, Amanda e Lucélia pelo incentivo, parceria e amizade. Foram longas noites de estudo para as avaliações, apresentações e escrita científica.

A Universidade Federal de Mato Grosso do Sul, ao Programa de Pós Graduação em Agronomia, pelas oportunidades e o espaço cedido para desenvolvimento desta pesquisa.

Todos os professores da universidade, o meu obrigado por todo conhecimento e dedicação com os alunos do curso de pós- graduação em Agronomia.

À minha orientadora Dra Ana Carina da Silva Cândido Seron, pela dedicação e atenção, pela paciência e oportunidades que me concedeu.

Aos membros da banca examinadora Dra. Mayara Fávero Cotrim e Dra. Dthenifer Cordeiro Santana, pela participação, correções e sugestões para melhoria deste trabalho.

Aos membros do laboratório de UPLC que auxiliaram durante a condução do experimento.

A todos, o meu muito obrigada!

## LISTA DE FIGURAS

|                                                                                                                                                                                                                                                      |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 1.</b> Temperatura e pluviosidade média ao longo do desenvolvimento do experimento.....                                                                                                                                                    | 10 |
| <b>Figura 2.</b> PCA-kmeans para a separação dos grupos de acordo com a similaridade do teor de aminoácidos e o comportamento espectral dos materiais de soja.....                                                                                   | 14 |
| <b>Figura 3.</b> comparação de médias para os teores de ácido aspártico, ácido glutâmico alanina, arginina, cistina, fenilalanina, glicina e histidina para cada grupo formado. Letras distintas diferem pelo teste de SK a 5% de significância..... | 15 |
| <b>Figura 4.</b> Comparação de médias para os teores de isoleucina, leucina, lisina, metionina, prolina, serina, tirosina, treonina e valina para cada grupo formado. Letras distintas diferem pelo teste de SK a 5% de significância.....           | 16 |
| <b>Figura 5.</b> Comparação de médias para as bandas B1-B14 para cada grupo formado. Letras distintas diferem pelo teste de SK a 5% de significância.....                                                                                            | 17 |
| <b>Figura 6.</b> Comparação de médias para as bandas B15-B28 para cada grupo formado. Letras distintas diferem pelo teste de SK a 5% de significância.....                                                                                           | 17 |
| <b>Figura 7.</b> Correlação de Pearson entre os aminoácidos e o dados espectrais para cada grupo formado.....                                                                                                                                        | 20 |

## LISTA DE TABELA

|                                                                                                                                              |    |
|----------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Tabela 1.</b> Intervalos de faixa espectral (nm) e comprimento de onda médio (nm) usados no estabelecimento das 28 bandas espectrais..... | 11 |
|----------------------------------------------------------------------------------------------------------------------------------------------|----|



# RELAÇÃO ENTRE DADOS HIPERESPECTRAIS E A COMPOSIÇÃO DE AMINOÁCIDOS EM GENÓTIPOS DE SOJA

**RESUMO:** A refletância espectral de plantas pode ser prontamente relacionada aos parâmetros fisiológicos e bioquímicos. Assim, relacionar dados espectrais com teores de AA em diferentes materiais genéticos fornece uma abordagem inovadora e eficiente para entender e manejar a diversidade genética. O objetivo do trabalho: (I) separar materiais genéticos de acordo com o teor de aminoácido e a refletância espectral e (II) estabelecer relação entre aminoácidos e as bandas espectrais calculadas a partir de dados hiperespectrais. Na safra 2023/2024, com 32 genótipos de soja, em blocos casualizados com quatro repetições. A análise espectral foi realizada 60 dias após a emergência (DAE). As amostras foram levadas ao laboratório onde efetuou-se a leitura da refletância com espectrorradiômetro, que tem a capacidade de medir o espectro na faixa de 350 a 2500 nm. Os comprimentos de onda, foram agrupados em médias de intervalos representativos, organizados em 28 bandas. Inicialmente, os genótipos foram submetidos à análise de agrupamento por meio do algoritmo k-means. No presente estudo, o valor de k que melhor discriminou os grupos de acordo com as características foi estabelecido em 4. Foi realizado uma correlação afim de averiguar a relação entre as variáveis dentro de cada grupo. As médias dos resultados foram agrupadas através do teste de Scott-Knott. A refletância das folhas de soja foi eficiente para separar os genótipos de soja de acordo com o teor de aminoácidos da folha. As bandas espectrais apresentam relações com os aminoácidos. A banda B21 a B28 apresenta maiores relações com positivas com ácido glutâmico, arginina, cistina, isoleucina, leucina, metionina, prolina, treonina e valina. A B12 a B17 relacionam positivamente com ácido glutâmico, histidina, metionina, prolina e treonina. B23 e B26 apresentam correlações positivas com arginina, cistina, prolina, tirosina e valina. De forma negativa: B2 a B10 com ácido aspártico, alanina, arginina, metionina, prolina, serina, tirosina e valina. B15 a B18 com ácido aspártico, alanina, cistina, metionina, prolina e serina e B27 com glicina, histidina, lisina, metionina, prolina, tirosina e valina.

**Palavras-chave:** Sensoriamento remoto. Fenotipagem de alta de precisão. Proteína.

## ABSTRACT

The spectral reflectance of plants can be readily correlated with physiological and biochemical parameters. Thus, relating spectral data to amino acid (AA) content in different genetic materials provides an innovative and efficient approach to understanding and managing genetic diversity. The objective of this work was: (I) to separate genetic materials according to amino acid content and spectral reflectance and (II) to establish a relationship between amino acids and spectral bands calculated from hyperspectral data. In the 2023/2024 growing season, 32 soybean genotypes were used in randomized blocks with four replications. Spectral analysis was performed 60 days after emergence (DAE). The samples were taken to the laboratory where reflectance was measured using a spectroradiometer, which has the capacity to measure the spectrum in the range of 350 to 2500 nm. The wavelengths were grouped into averages of representative intervals, organized into 28 bands. Initially, the genotypes were subjected to cluster analysis using the k-means algorithm. In the present study, the k value that best discriminated the groups according to the characteristics was set at 4. A correlation was performed to ascertain the relationship between the variables within each group. The averages of the results were grouped using the Scott-Knott test. Soybean leaf reflectance was efficient in separating soybean genotypes according to leaf amino acid content. The spectral bands show relationships with amino acids. Bands B21 to B28 show the strongest positive correlations with glutamic acid, arginine, cystine, isoleucine, leucine, methionine, proline, threonine, and valine. Bands B12 to B17 are positively correlated with glutamic acid, histidine, methionine, proline, and threonine. Bands B23 and B26 show positive correlations with arginine, cystine, proline, tyrosine, and valine. Vitamins B12 to B17 are positively correlated with glutamic acid, histidine, methionine, proline, and threonine. Vitamins B23 and B26 show positive correlations with arginine, cystine, proline, tyrosine, and valine. Negatively correlated: Vitamins B2 to B10 are correlated with aspartic acid, alanine, arginine, methionine, proline, serine, tyrosine, and valine; Vitamins B15 to B18 are correlated with aspartic acid, alanine, cystine, methionine, proline, and serine; and Vitamin B27 is correlated with glycine, histidine, lysine, methionine, proline, tyrosine, and valine.

**KEYWORDS:** Remote sensing. High-precision phenotyping. Protein.

## SUMÁRIO

|                          |    |
|--------------------------|----|
| INTRODUÇÃO .....         | 8  |
| MATERIAL E MÉTODOS ..... | 9  |
| RESULTADOS .....         | 13 |
| DISCUSSÃO.....           | 21 |
| CONCLUSÃO .....          | 23 |
| REFERENCIAS .....        | 24 |

## INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) é uma importante fonte de proteína e óleo vegetal para animais e humanos, possuindo uma ampla gama de usos em todo o mundo em forma de alimentos e ração (Savary et al., 2019). Possui diversos aminoácidos (AA) essenciais em sua composição (Kanchana et al., 2015), todavia varia conforme a característica genética do material cultivado. Portanto, tornou-se importante em programas de melhoramento de plantas selecionar genótipos de soja quanto a suas qualidades nutricionais, principalmente em relação aos teores de aminoácidos essenciais. Por exemplo, a soja transgênica com O-acetilserina sulfidrilase (OASS) foi desenvolvida para o aumento do conteúdo de cisteína (Kanchana et al., 2016; Kim et al., 2012), todavia há outros aminoácidos que podem ser aumentados na espécie.

A espectrofotometria, a análise química, espectrometria de massa e a cromatografia líquida são os principais métodos para determinar o conteúdo de aminoácidos, pois esses métodos podem estimar uma variedade de aminoácidos e têm as vantagens de alta sensibilidade e precisão (Shu et al., 2022). Além disso, a utilização de tecnologia de dados hiperespectral pode fornecer informações importantes e complementares relacionadas a parâmetros fisiológicos e bioquímicos das culturas, com as vantagens de ser um método de análise rápido, de alto rendimento e não destrutivo (Z. Li et al., 2019; Mao et al., 2020), podendo ser um forte aliado na determinação de aminoácidos em culturas de interesse.

A utilização de sensores hiperespectrais é uma tecnologia não destrutiva e simples de operar que é rápida na análise, que permite a mensuração de várias amostras simultaneamente e pode remediar efetivamente algumas deficiências dos métodos tradicionais (Liang et al., 2023). Sensores abrangentes da região do visível (VIS), infravermelho próximo (NIR) e infravermelho próximo de onda curta (SWIR) fornecem informações com variações na resolução espacial e espectral disponibilizando informações mais abrangentes, levando a muitos dados de alta dimensão espectral (Long et al., 2024).

Silva Junior et al. (2018), propôs reduzir dados hiperespectrais em 28 bandas, que apresentam as médias de intervalos significativos de comprimento de onda e pontos de inflexão que são representativos de elementos conhecidos como por exemplo a e b caroteno, xantofila, clorofila e regiões do espectro relacionadas com absorção da clorofila e água.

Os mesmos autores destacaram, na faixa de comprimentos de onda entre 400 e 700 nm, correspondente à região do visível (VIS), a refletância foi considerada menor, devido à absorção de luz pelos pigmentos presentes nas folhas, especialmente a clorofila.

Um leve aumento na reflectância é observado em torno de 550 nm, relacionado à predominância do verde no espectro visível (Silva Junior et al., 2018). Na região do infravermelho próximo (NIR), que abrange de 700 a 1300 nm, a reflectância apresenta aumento significativo, influenciado pela estrutura interna das folhas, incluindo o formato e o tamanho das células, bem como a proporção de espaços intercelulares. Já na faixa de 1300 a 2500 nm, correspondente ao infravermelho de ondas curtas (SWIR), ocorre uma redução gradativa da reflectância, refletindo as características de absorção relacionadas ao conteúdo de água líquida nas folhas (Silva Junior et al., 2018).

Portanto, a reflectância espectral pode ser prontamente relacionada aos parâmetros fisiológicos e bioquímicos de culturas agrícolas (Pandey e Prabhakar, 2016). Assim, relacionar dados espectrais com teores de AA em diferentes materiais genéticos fornece uma abordagem inovadora e eficiente para entender e manejar a diversidade genética. Além disso, o uso dessa tecnologia irá otimizar o manejo agrícola e acelerar programas de melhoramento genéticos de planta, identificando genótipos com características desejáveis sem a necessidade de análises químicas detalhadas em todas as amostras. Essa integração de dados bioquímicos e espectrais representa uma ferramenta poderosa para a agricultura de precisão contribuindo para a sustentabilidade agrícola.

A partir dessa perspectiva, relacionar dados espectrais com teores de AA pode fornecer uma abordagem moderna e eficaz para explorar a diversidade genética, melhorar a gestão agrícola e contribuir com programas de melhoramento genético de plantas fornecendo uma abordagem rápida, não destrutiva e informativa para avaliar características fenotípicas complexas, como o teor de aminoácidos, em diferentes materiais genéticos e possibilitando avaliar grandes populações de plantas simultaneamente, reduzindo o tempo e os recursos necessários para a seleção. Assim, objetivou-se com o trabalho: (I) separar materiais genéticos de acordo com o teor de aminoácido e a reflectância espectral e (II) estabelecer relação entre aminoácidos e as bandas espectrais calculadas a partir de dados hiperespectrais.

## **MATERIAL E MÉTODOS**

O experimento foi conduzido em campo experimental localizado na Universidade Federal de Mato Grosso do Sul, município de Chapadão do Sul (18°41'33''S, 52°40'45''W, com 810 m de altitude média), Mato Grosso do Sul, Brasil. na safra 2023/2024, com 32 genótipos de soja, em blocos casualizados com quatro repetições. Cada parcela experimental foi constituída por quatro linhas de um metro de comprimento com espaçamento de 0,45 m e

densidade de 15 sementes por metro. O clima da região é Tropical de Savana (Aw), segundo a classificação de Köppen e Geiger, a distribuição pluviométrica e a temperatura média ao longo do ciclo consta na Figura 1. O solo foi classificado como Latossolo Vermelho Distrófico Argiloso, apresentando as seguintes características na camada de 0 – 0,20 m: pH (H<sub>2</sub>O) = 6,2; Al trocável (cmol<sub>c</sub> dm<sup>-3</sup>) = 0,0; Ca+Mg (cmol<sub>c</sub> dm<sup>-3</sup>) = 4,31; P (mg dm<sup>-3</sup>) = 41,3; K (cmol<sub>c</sub> dm<sup>-3</sup>) = 0,2; Matéria orgânica (g dm<sup>-3</sup>) = 19,74; V (%) = 45; m (%) = 0,0; Soma de bases (cmol<sub>c</sub> dm<sup>-3</sup>) = 2,3; CTC (cmol<sub>c</sub> dm<sup>-3</sup>) = 5,1.

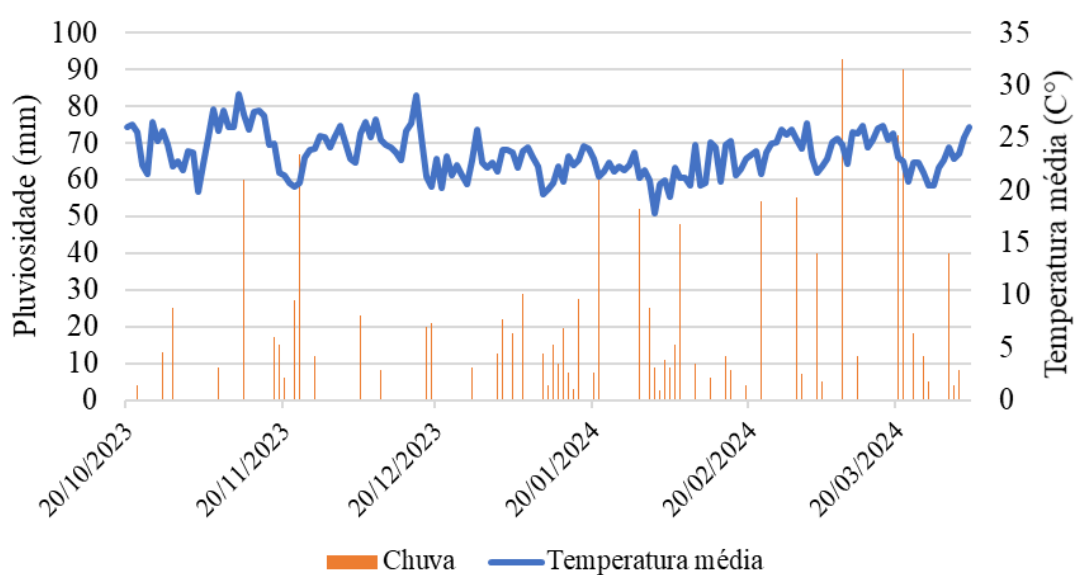


Figura 1. Temperatura e pluviosidade média ao longo do desenvolvimento do experimento

Antes da semeadura realizou-se o preparo do solo de maneira convencional, por meio de aração e gradagem niveladora. As sementes dos genótipos de soja foram tratadas com mistura de fungicida (Piraclostrobina + Metil Tiofanato) e inseticida (Fipronil), os quais foram aplicados na proporção de 200 mL do produto para cada 100 kg de sementes. Além disso, realizou-se a inoculação com *Bradyrhizobium* diretamente no sulco de semeadura, na dose indicada pelo fabricante. Durante todo o ciclo da cultura, foram realizados tratos culturais com o uso de fungicidas, inseticidas e herbicidas conforme a necessidade da cultura.

A análise espectral foi realizada 60 dias após a emergência (DAE), onde foram coletadas três amostras foliares do terceiro trifólio, de cima para baixo, bem desenvolvido, de cada parcela. As amostras foram levadas ao laboratório onde efetuou-se a leitura da reflectância com espectrorradiômetro (FieldSpec 4 Jr, Analytical Spectral Devices, Boulder, EUA), que tem a capacidade de medir o espectro na faixa de 350 a 2500 nm. Para cada leitura da folha, foi utilizado o equipamento ASD Plant Probe. Na calibração do equipamento utilizou-se uma placa

branca, de sulfato de bário, que reflete 100% da luz. O equipamento foi calibrado ao final de cada leitura das amostras de cada bloco, totalizando quatro calibrações. O sensor conectado a um computador registrou cada leitura por meio de um software próprio do sensor chamado RS<sup>3</sup>. Em seguida, os arquivos foram importados para outro software, ViewSpectroPro para extração dos dados. Os comprimentos de onda, foram agrupados em médias de intervalos representativos, organizados em bandas (Tabela 1) (Silva Junior et al. 2018). Estas faixas estão associadas com processos fisiológicos importantes para a planta, como foi descrito por Silva Junior et al. (2018).

Tabela 1. Intervalos de faixa espectral (nm) e comprimento de onda médio (nm) usados no estabelecimento das 28 bandas espectrais.

| <b>Banda (nº)</b> | <b>Faixa espectral (nm)</b> | <b>Comprimento de onda médio (nm)</b> |
|-------------------|-----------------------------|---------------------------------------|
| 1                 | 350–369                     | 359.5                                 |
| 2                 | 370                         | 370.0                                 |
| 3                 | 371–419                     | 395.0                                 |
| 4                 | 420                         | 420.0                                 |
| 5                 | 421–424                     | 422.5                                 |
| 6                 | 425                         | 425.0                                 |
| 7                 | 426–444                     | 435.0                                 |
| 8                 | 445–475                     | 460.0                                 |
| 9                 | 480                         | 480.0                                 |
| 10                | 481–500                     | 490.5                                 |
| 11                | 501–530                     | 515.5                                 |
| 12                | 531–539                     | 535.0                                 |
| 13                | 540                         | 540.0                                 |
| 14                | 541–649                     | 595.0                                 |
| 15                | 650                         | 650.0                                 |
| 16                | 661–670                     | 665.5                                 |
| 17                | 675                         | 675.0                                 |
| 18                | 676–684                     | 680.0                                 |
| 19                | 685–689                     | 687.0                                 |
| 20                | 690–700                     | 695.0                                 |
| 21                | 701–709                     | 705.0                                 |

| <b>Banda (n°)</b> | <b>Faixa espectral (nm)</b> | <b>Comprimento de onda médio (nm)</b> |
|-------------------|-----------------------------|---------------------------------------|
| 22                | 710                         | 710.0                                 |
| 23                | 711–730                     | 720.5                                 |
| 24                | 960                         | 960.0                                 |
| 25                | 1100                        | 1100.0                                |
| 26                | 1400                        | 1400.0                                |
| 27                | 1930                        | 1930.0                                |
| 28                | 2200                        | 2200.0                                |

Após as análises as amostras foliares de cada parcela foram secas em estufas com circulação forçada de ar a 65° C e posteriormente moídas para a realização das análises de quantificação de aminoácidos. Para a análise dos aminoácidos L-histidina, L-serina, L-arginina, L-glicina, L-ácido aspártico, L-ácido glutâmico, L-treonina, L-alanina, L-prolina, L-cistina, L-lisina, L-tirosina, L-metionina, L-valina, L-isoleucina, L-leucina e L-felilalanina o material vegetal foi submetido a hidrólise ácida. Para a extração dos aminoácidos 50 mg das amostras secas e moídas foram submetidas ao processo de digestão ácida com ácido clorídrico a 6 N em blocos de digestão a 110° C por 24 horas. Após esse período 1 mL da solução foi filtrada por meio de uma seringa com um filtro de 0,2 µm e transferidos para tubos de ensaio. Posteriormente, uma alíquota de 50 µL foi transferida para um vial e mantido em dessecador com sílica gel e sob vácuo por aproximadamente 16 horas, ou até completa secura do solvente.

Após esse período, foi adicionado ao resíduo seco 20 µL do diluente de amostra (HCl 20mM) agitando vigorosamente em vórtex durante 15 segundos. Posteriormente foi adicionado 60 µL de tampão borato e o vial foi novamente agitado em vórtex durante 15 segundos. Por fim, foram adicionados 20 µL do derivatizante (6-aminoquinolil-N-hidroxisuccinimidil carbamato - AQC) adquirido na forma de Kit (Waters®) e a mistura foi agitada imediatamente em vórtex durante 10 segundos. O vial foi deixado em repouso durante 1 minuto para o término da reação. Posteriormente foi transferido para banho-maria a 55 ± 1° C por 10 minutos. A solução assim obtida foi levada para análise cromatográfica. Alíquotas de 1 µL foram utilizadas para injeção direta no equipamento. Cada amostra teve três repetições de análise

A separação e quantificação dos 17 aminoácidos foram realizadas em cromatógrafo líquido de ultra performance (UPLC) Waters Acquity série 1100, com injetor automático de amostras. Foi utilizado uma coluna de fase reversa do tipo C18, 1,8 µm (diâmetro interno de 2,1 mm (i.d.) £ 100 mm. Para a separação foi adotado o sistema de gradiente linear com quatro fases móveis, tendo como fases móvel: A (100% AccQTag Ultra eluente A concentrado), B



(90:10 Água: AccQTag Ultra eluente B), C (100% água MilliQ), D (100% AccQ•Tag Ultra eluent B). A vazão da fase móvel foi de 0,7 mL min<sup>-1</sup> e a temperatura da coluna durante a corrida foi de 49°C. A detecção dos aminoácidos foi realizada utilizando detector de fluorescência, ajustado para o comprimento de onda de 266 a 473 nm (GARZA-ULLOA et al., 1986). A identidade qualitativa e quantitativa do pico foi confirmada comparando os tempos de retenção e os espectros de UV de compostos individuais e pelo método de adição padrão.

Inicialmente, os genótipos foram submetidos à análise de agrupamento por meio do algoritmo k-means. Primeiramente, o algoritmo k-means é empregado para calcular as distâncias entre cada amostra e os centroides dentro de cada classe. Com base nessas distâncias, as amostras são atribuídas a seus respectivos grupos, considerando-as semelhantes às amostras pertencentes ao mesmo grupo e distintas das do outro grupo. Os clusters resultantes do k-means são então representados em um gráfico biplot por meio da análise de componentes principais (PCA), ilustrando a separação entre eles. Esse método requer a definição prévia do valor de k, que representa o número de grupos desejado antes da execução da análise (AHMED et al., 2020).

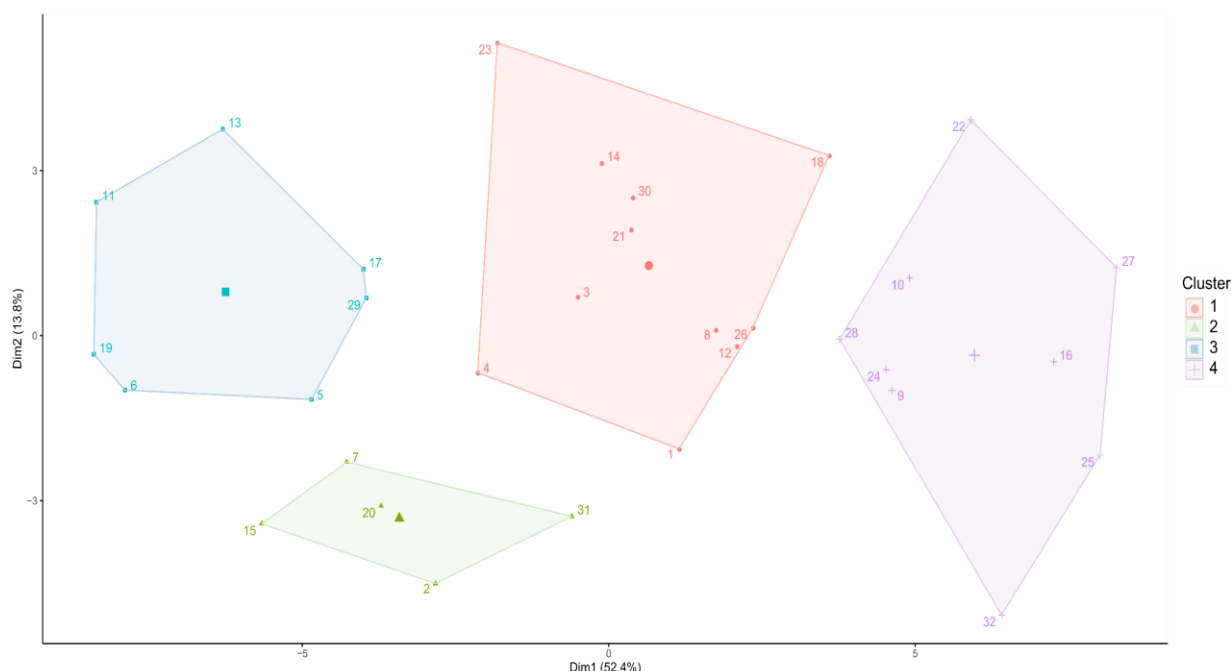
No presente estudo, o valor de k que melhor discriminou os grupos de acordo com as características foi estabelecido em 4, resultando na separação dos genótipos em quatro grupos distintos entre eles. Foram criados boxplots comparativos para analisar as médias de cada AA e bandas, com o objetivo de observar as diferenças entre os grupos formados. Foi realizado uma correlação afim de averiguar a relação entre as variáveis dentro de cada grupo, representado por uma rede de correlação em que as linhas vermelhas representam correlações negativas e as linhas verdes correlações positivas, as linhas destacadas de maior espessura são correlações acima de 0.6 estabelecida pelo próprio programa em que o gráfico foi confeccionado (Rbio, Bhering, 2017). Devido as baixas correlações entre os AA e as bandas foi estabelecido correlações significativas acima e ressaltas acima de 0.20.

As médias dos resultados foram agrupadas através do teste de Scott-Knott (Scott; Knott, 1974) com nível de significância de 5%. Para a análise de PCA-Kmeans e os demais gráficos e realização dos testes de médias, foram empregados o software R (Team, 2013) e os pacotes ggplot2, ExpDes.pt e Metan.

## RESULTADOS

Os 32 materiais de soja foram submetidos a análise de PCA-Kmeans para separar os genótipos de acordo com os teores de aminoácidos e o comportamento espectral. Dessa forma

a utilização da análise multivariada permitiu a separação em quatro grupos distintos (Figura 1). A maior quantidade de materiais ficou agrupados no grupo 1 (Cluster 1), totalizando 11 genótipos e no grupo 2 (Cluster 2) se concentrou o menor número de genótipos, cinco materiais.



**Figura 2.** PCA-kmeans para a separação dos grupos de acordo com a similaridade do teor de aminoácidos e o comportamento espectral dos materiais de soja

Os teores de ácido aspártico, ácido glutâmico, alanina, arginina, cistina, fenilalanina, glicina e histidina foram comparados segundo o agrupamento de Scott-Knott para observar as maiores e menores teores de cada aminoácido dentro de cada grupo. O C1 alcançou maiores teores de ácido aspártico, ácido glutâmico, alanina, cistina e fenilalanina. O C2 teve como característica as menores concentrações de ácido aspártico, ácido glutâmico, alanina, cistina e fenilalanina e obteve maiores teores de arginina e glicina. O grupo C3 apresentou teores medianos para o ácido aspártico, ácido glutâmico alanina, arginina, cistina, fenilalanina, glicina e histidina. C4 apresentou maiores teores apenas dos aminoácidos ácido aspártico, alanina e cistina. Histina apresentou teores similares entre todos os grupos.

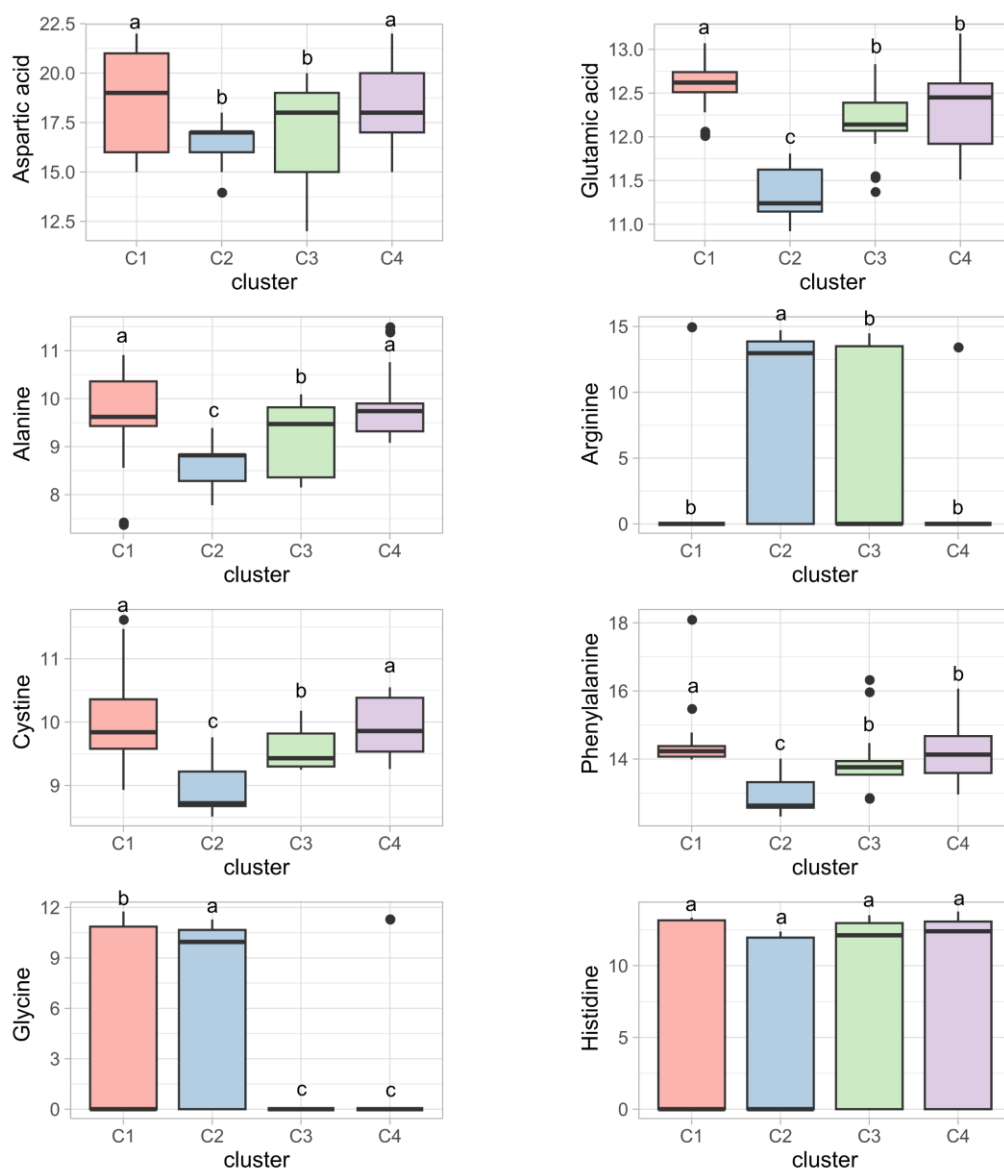


Figura 3. comparação de médias para os teores de ácido aspártico, ácido glutâmico alanina, arginina, cistina, fenilalanina, glicina e histidina para cada grupo formado. Letras distintas diferem pelo teste de SK a 5% de significância.

Os aminoácidos isoleucina, leucina, lisina, prolina, tirosina, treonina e valina apresentaram maiores teores nos genótipos C1. C2 apresentou os menores teores de isoleucina, leucina, lisina, prolina, tirosina e valina, e apresentou teores elevados de metionina. C3 apresentou teores medianos de isoleucina, leucina, prolina, serina, treonina e valina; e apresentou altos teores de lisina e tirosina. C4 apresentou valores elevados de lisina, metionina, tirosina e treonina, sendo valores medianos de isoleucina, leucina, prolina e valina.

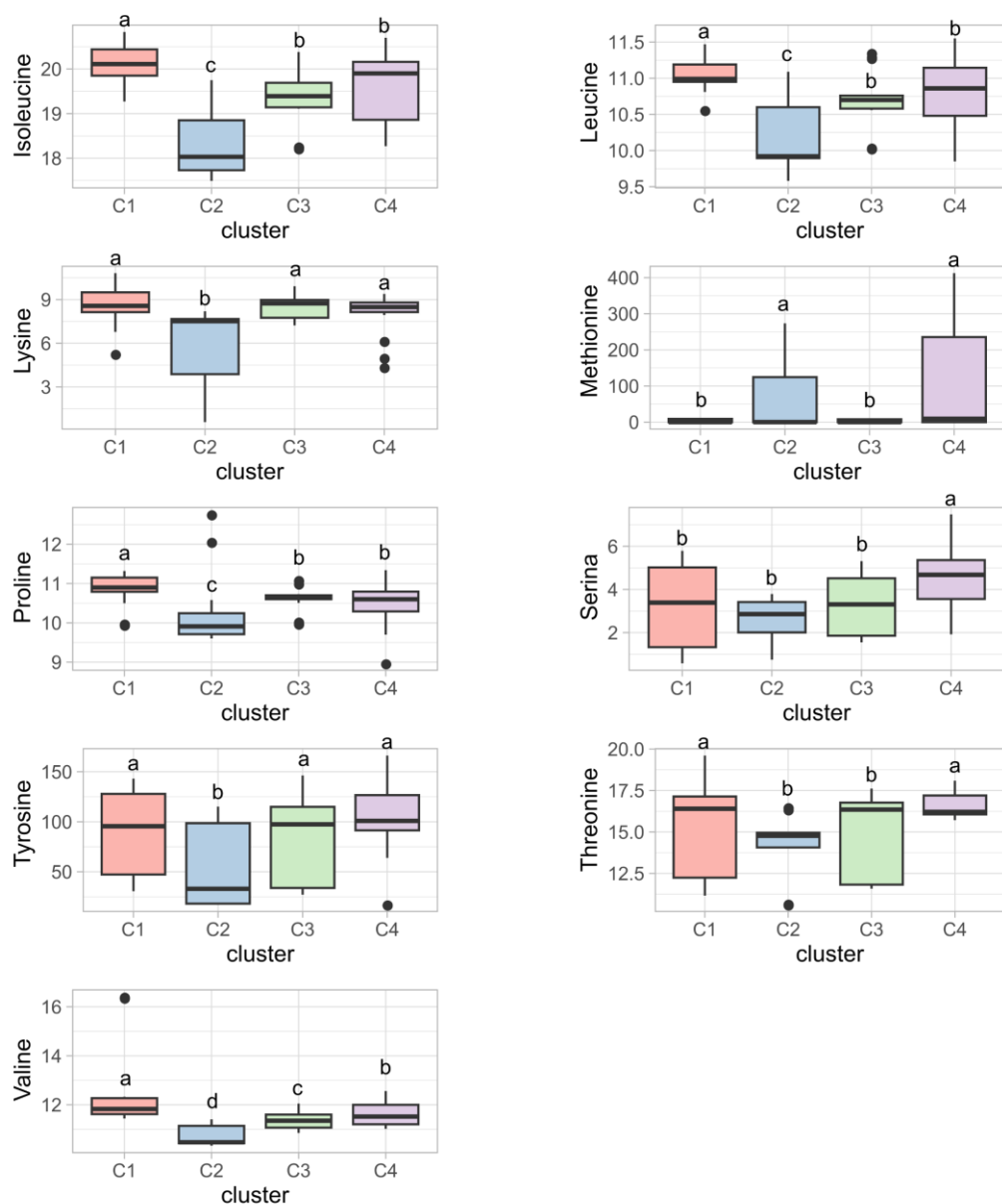


Figura 4. Comparação de médias para os teores de isoleucina, leucina, lisina, metionina, prolina, serina, tirosina, treonina e valina para cada grupo formado. Letras distintas diferem pelo teste de SK a 5% de significância.

Quanto as bandas espectrais houve um comportamento específico para cada grupo entre as bandas B1-B14, em que C4 teve maior refletância em todas as bandas, seguido do C1. De uma forma geral C3 teve a menor refletância, não havendo diferença de C2 entre as bandas B7-B10. Os grupos C1 e C2 tiveram comportamentos similares entre as bandas B12 e B13.

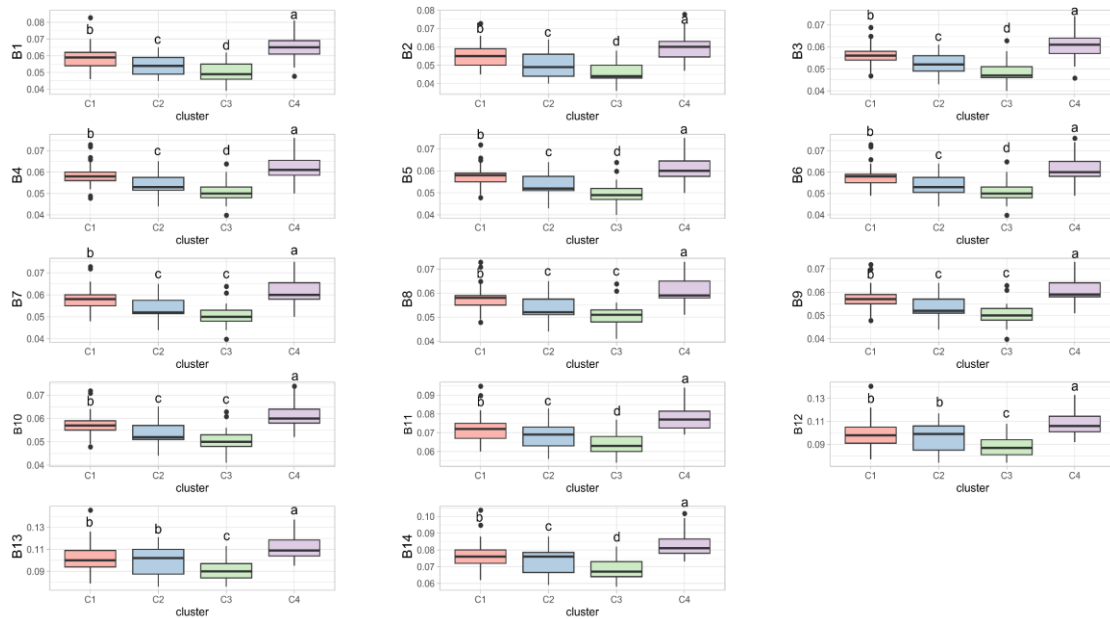


Figure 5. Comparação de médias para as bandas B1-B14 para cada grupo formado. Letras distintas diferem pelo teste de SK a 5% de significância.

Quanto as bandas B15-B28 o cluster C4 permaneceu com a maior reflectância, seguido de C1, C3 e C2. Entre B17-B20 o comportamento espectral entre C2 e C3. B21-B26 e B28 tiveram o mesmo comportamento espectral entre C1 e C2.

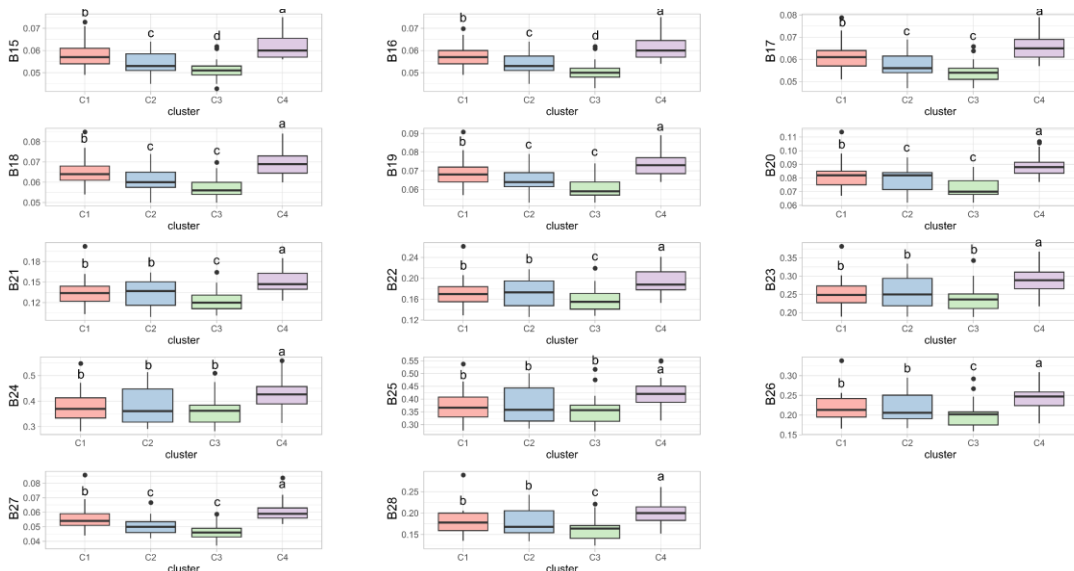


Figure 6. Comparação de médias para as bandas B15-B28 para cada grupo formado. Letras distintas diferem pelo teste de SK a 5% de significância.

A correlação de Pearson teve comportamentos distintos dentro de cada grupo entre os aminoácidos e as bandas espectrais (Figura 6). A correlação entre bandas foi positiva entre elas em todos os grupos. No grupo 1 arginina e prolina teve alta correlação negativa (-0.81); houve alta correlação positiva ( $>0.60$ ) entre os aminoácidos serina, alanina, tirosina, treonina, cistina, isoleucina, glutamina, lisina. A correlação entre as bandas e os aminoácidos no grupo tiveram alguns resultados com correlações acima de -0.20 que foram interessantes: fenilalanina teve correlações negativas acima de 0.20 chegando a -0.36 com todas as 28 bandas; ácido aspártico teve correlações negativas acima de -0.20 com todas as bandas exceto B4 e B5; arginina teve correlação negativa  $>-0.20$  com as bandas B21-B26 e B28; glicina teve uma correlação de -0.21 com B27; lisina com B7-10 correlações negativas de -0.20; serina e tirosina tiveram correlações negativas de -0.20 com B5.

No grupo 2 houve correlações maiores e com mais aminoácidos do que no grupo 1. Ácido glutâmico teve correlações em torno de 0.20 com as bandas B12-B14 e B22-B23; alanina teve correlações negativas variando entre -0.22 e -0.51 com todas as bandas, havendo a maior correlação com B9 (-0.51); arginina alcançou correlações de -0.20 com as bandas B16 e B17; cistina com B27 e B28, correlações positivas 0.20; histina teve correlações positivas com B11-B14 e B20-B26 variando entre 0.21 e 0.39 e correlação negativa de -0.21 com B27; lisina teve correlações negativas com B5, B6-B10 e B27; metionina alcançou correlações positivas entre 0.21 e 0.36 com: B1, B11-B14, B19-B26 e B28; Prolina também alcançou correlações positivas entre 0.21 e 0.36 com B1-B11, B14-B18, B24, B26-B28; serina teve correlações negativas em torno de -0.20 com B5, B7-B10; tirosina teve correlações negativas em torno de -0.20 com B11-B14 e valina teve correlação de 0.24 com B28.

No grupo 3, as correlações entre aminoácidos (AA): ácido aspártico teve alta correlação positiva ( $>0.60$ ) com alanina, serina, tirosina e treonina; ácido glutâmico teve alta correlação positiva com fenilalanina, isoleucina, leucina, prolina e valina; alanina teve alta correlação positiva com ácido aspártico, cistina, serina, tirosina e treonina; arginina teve correlação positiva com tirosina. Cistina teve alta correlação positiva com alanina, histidina e valina. Fenilalanina teve alta correlação positiva com ácido glutâmico, isoleucina, leucina, prolina e valina; histidina teve alta correlação apenas com cistina; isoleucina, leucina e prolina tiveram alta correlação positiva com ácido glutâmico, fenilalanina, prolina e valina. Serina teve alta correlação positiva com ácido aspártico, alanina, tirosina e treonina. Tirosina teve alta correlação com ácido aspártico, alanina, arginina, histidina, treonina e serina; treonina teve correlação positiva com ácido aspártico, serina, tirosina e alanina.

Já a Correlação entre aminoácidos e bandas espectrais: nota-se que ácido aspártico teve correlação negativa ( $>-0,25$ ) com B2-B11 e B15-B18; ácido glutâmico teve correlação positiva ( $>0,20$ ) com B2-B10, B15-B16 e B7. Alanina correlacionou-se de forma negativa ( $>-0,25$ ) com B2-B10 e B15-B17; positiva com B28; arginina teve correlação negativa ( $-0,21$ ) com B15; positiva ( $>0,20$ ) com B23, B26 e B28. Histidina teve correlação positiva ( $0,21$ ) com B1. Isoleucina teve correlação positiva ( $>0,25$ ) com B2-B4, B6-B10, B15-B16. Metionina, negativa ( $-0,26$  a  $-0,38$ ) com B1-B10, B15-B19 e B27; prolina, positiva com B2, B4, B6-B8 e B10; serina, negativa ( $-0,21$  a  $-0,48$ ) com B2-B3, B5-B11, B15-B18 e B27; positiva com B22-B23; tirosina, negativa com B2, B4, B7-B10, B15-B16; positiva com B21-B23, B26 e B28; valina, positiva ( $0,28$ ) com B2.

No grupo 4 houve correlações altas e positivas ( $>0,8$ ) entre alanina e ácido aspártico; ácido glutâmico com isoleucina e leucina, prolina, treonina e valina; alanina alcançou alta correlação com valina, tirosina, serina; arginina teve alta correlação negativa com isoleucina, leucina, tirosina e prolina e positiva de alta magnitude com metionina; ácido glutâmico, cistina alcançou alta correlação positiva com isoleucina, leucina, treonina e valina; fenilalanina obteve apenas correlação positiva de alta magnitude com leucina ( $0,65$ ). Glicina teve alta correlação negativa com serina ( $-0,61$ ); histina correlação positiva e de alta magnitude com tirosina e treonina; isoleucina alcançou alta correlação positiva com ácido glutâmico, cistina, leucina, prolina, treonina e valina; leucina teve alta correlação positiva com ácido glutâmico, cistina, fenilalanina, isoleucina, prolina, treonina e valina e negativa com arginina; metionina teve alta correlação positiva com arginina e negativa com tirosina; prolina teve alta correlação positiva com ácido glutâmico, isoleucina, leucina e negativa com arginina; serina teve altas correlações com ácido aspártico, alanina e negativa com glicina; tirosina teve alta correlação positiva com alanina, histidina e negativa com arginina; treonina teve alta correlação com ácido glutâmico, isoleucina, leucina e valina; valina teve altas correlações com ácido glutâmico, alanina, cistina, isoleucina, e treonina.

Quanto as correlações entre as bandas e os aminoácidos o ácido glutâmico teve correlações acima de  $-0,20$  com as bandas B2, B5, B7-B10, B14-B17, B22-B28; arginina teve correlação positiva acima de  $0,20$  com B14-B15 e B27 e negativa com B26; cistina teve correlação negativa com B2, B8-B10, B15-B17 e B27, e positiva com B21 e B23, B26 e B28; glicina teve correlação de  $-0,33$  com B27; isoleucina e leucina teve correlação negativa com B2, B8-B10, B15-B17 e positiva com as bandas B22-B28; lisina teve em torno de  $-0,20$  com B23-B25; metionina teve correlação positiva com B12-B14 e B21-B23; prolina teve correlação negativa com B2, B3-B4, B7-B10, B15-B17 e B27, e positiva com B23-B26 e B28; tirosina

teve correlação de -0.30 com B27; treonina teve correlação negativa com B2, B3-B4, B7-B10, B15-B16, B27 e positiva com B12, B13, B21-B26 e B28; valina teve correlação negativa em torno -0.20 com as bandas B8-B10, B15-B16 e B27 e positiva com B21-B26 e B28

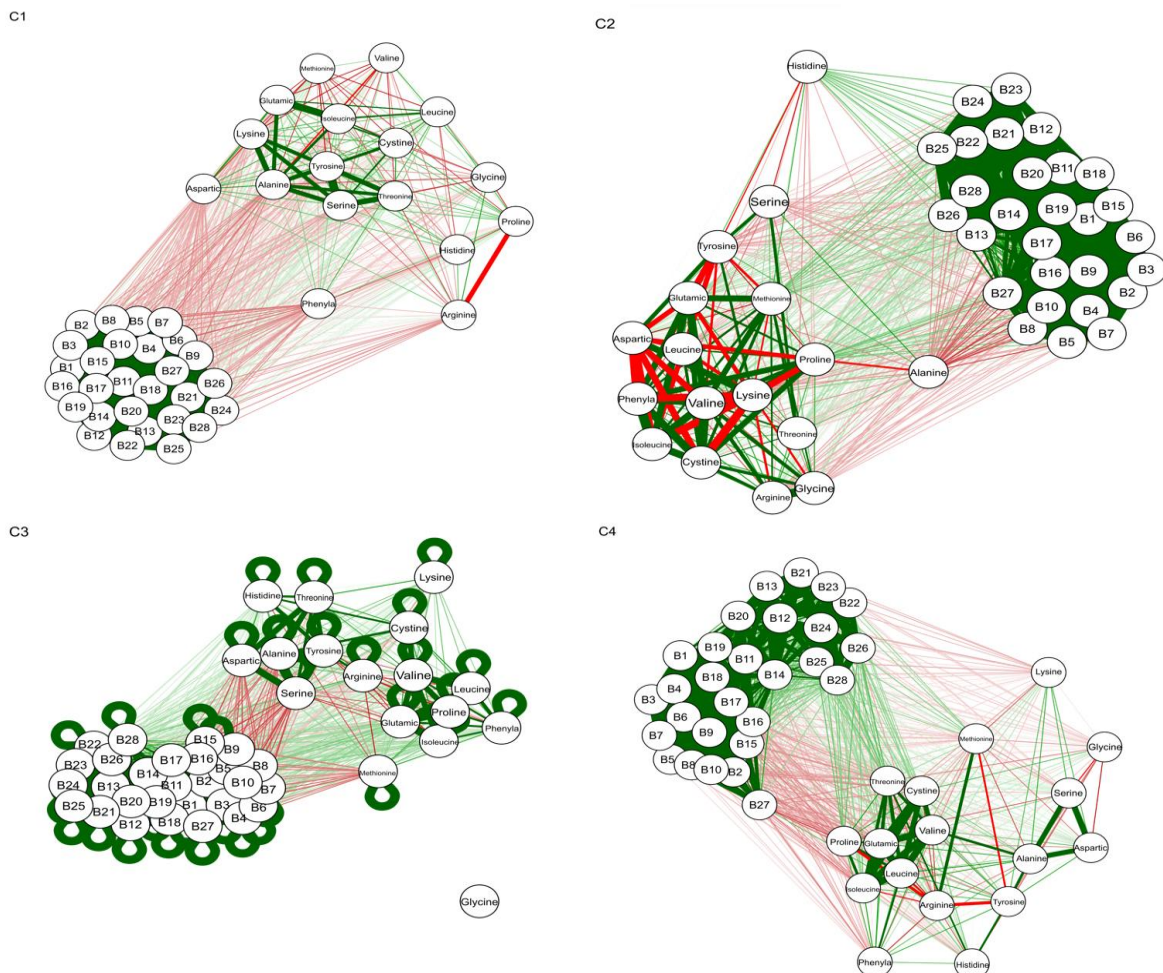


Figura 7. Correlação de Pearson entre os aminoácidos e o dados espectrais para cada grupo formado.

Os grupos formados pelos diferentes materiais analisados apresentaram perfis distintos tanto em composição de aminoácidos quanto no comportamento espectral. De forma geral os materiais pertencentes ao grupo 1 se destacaram por apresentar os maiores teores da maioria dos aminoácidos, como ácido aspártico, ácido glutâmico, alanina, cistina, fenilalanina, isoleucina, leucina, lisina, prolina, tirosina, treonina e valina, indicando superioridade em termos de composição de aminoácidos. O grupo 2 caracterizou-se pelas menores concentrações da maioria dos aminoácidos, exceto por apresentar teores elevados de arginina, glicina e metionina. Já o grupo 3 apresentou teores intermediários para a maioria dos aminoácidos, com



exceção de altos teores de lisina e tirosina. E o grupo 4 destacou-se por maiores teores de lisina, metionina, tirosina e treonina, com valores medianos para isoleucina, leucina, prolina e valina. Quanto ao comportamento espectral o C4 apresentou maior refletância em todas as bandas espectrais (B1-B28), seguido do C1. O C3 demonstrou a menor refletância na maioria das bandas, com comportamento similar ao C2 nas bandas B7-B10, e ao C1 nas bandas B21-B26 e B28. O C1 e o C2 exibiram comportamentos semelhantes nas bandas B12 e B13.

Quanto as correlações, as bandas que apresentaram mais relações com os aminoácidos, considerando tanto correlações positivas foram: B21 a B28 destacaram-se por correlações positivas com aminoácidos como ácido glutâmico, arginina, cistina, isoleucina, leucina, metionina, prolina, treonina e valina; B12 a B17: Relacionaram-se positivamente com ácido glutâmico, histidina, metionina, prolina e treonina; B23 e B26: Apresentaram correlações positivas consistentes com arginina, cistina, prolina, tirosina e valina. De forma negativa: B2 a B10: Foram mais frequentemente associadas a correlações negativas com ácido aspártico, alanina, arginina, metionina, prolina, serina, tirosina e valina; B15 a B18: Mostraram correlações negativas significativas com ácido aspártico, alanina, cistina, metionina, prolina e serina e B27: Destacou-se por correlações negativas com glicina, histidina, lisina, metionina, prolina, tirosina e valina.

## **DISCUSSÃO**

Segundo Guo et al. (2022), dentre os focos de maior importância na produção e pesquisa de soja está garantir de forma sustentável proteína de soja, o que pode ser alcançado por meio do melhoramento para variedades de alto rendimento e alta proteína usando os recursos genéticos da soja. Dessa forma, a separação dos grupos de acordo com a informação do conteúdo de aminoácidos (AA) dos materiais, permitiu separá-los em grupos, com destaque para os materiais genéticos pertencentes ao grupo 1, que neste estudo se apresentaram os mais promissores no fornecimento dos aminoácidos. Agrupar genótipos pela composição de AA permitiu identificar materiais genéticos com características desejáveis para diferentes finalidades, como produção de grãos ou uso forrageiro, direcionando programas de melhoramento para seleção de materiais. Importante destacar que todos os materiais genéticos tiveram o mesmo manejo ao longo de todo o ciclo, desde adubação até controle de pragas e doenças. Dessa forma, garantiu-se que os materiais pertencentes ao grupo 1 são superiores aos demais na quantidade de AA presente nas folhas desses genótipos. Os demais grupos

apresentaram destaque para diferentes AA, sendo assim seleciona-los de acordo com o interesse dos componentes específico que se deseja melhorar.

Dentre os AA de maior teor nos materiais do grupo 1 está prolina, sendo um aminoácido importante que está relacionado à tolerância ao estresse hídrico em plantas, funcionando como osmoprotetor por desempenhar um papel essencial como antioxidante (Gill, 1998; Teixeira et al., 2020). Dessa forma os teores de aminoácidos, especialmente de prolina, pode refletir na eficiência metabólica que pode estar relacionado ao desempenho produtivo e à resistência a estresses, principalmente abióticos.

Na literatura há pouca informação sobre a aplicação de dados hiperespectral a parâmetros bioquímicos em folhas de plantas, especialmente sobre a relação de dados hiperespectrais com o conteúdo de aminoácidos em folhas de soja. Em nosso estudo uma importante relação observada entre os aminoácidos e as bandas espectrais calculadas a partir dos dados hiperespectrais foi notada. Os teores de AA refletiram a saúde metabólica da planta e sua interação com fatores ambientais podendo indicar estresses, enquanto os dados espectrais captam variações morfofisiológicas, como mudança em pigmentos. Relacionar essas informações ofereceu entendimento integrado da fisiologia das plantas com as genéticas e características dos materiais.

De forma geral, as bandas B21-B28 tiveram maior número de correlações positivas relevantes com diferentes aminoácidos, enquanto as bandas B2-B10 e B15-B18 apresentaram mais correlações negativas. Isso indica que essas faixas espectrais possuem maior potencial para distinguir as concentrações de aminoácidos específicos nas amostras analisadas. As bandas B2-B10 abrangem de 370 a 500 nm, B15-B18 abrangem de 650-684 e de B21-B28 abrangem faixas específicas da região do SWIR de 701-730 (B21-B23) e as demais bandas abrangem especificamente as faixas: 960, 1100, 1400, 1930 e 2200 nm. Shu et al. (2022), afirmam que as bandas que possuem maior relação com a maioria dos aminoácidos em folhas de milho estavam concentradas principalmente nas faixas de 505.39–604.95 nm e 651.21–714.10 nm, o que pode ser causado principalmente pela influência de vários pigmentos especialmente o conteúdo de clorofila. Os autores ainda afirmam que possuem relativamente poucos estudos sobre a detecção não destrutiva do conteúdo de aminoácidos em folhas por espectroscopia espectral, e que é necessário realizar mais estudos para provar a viabilidade da detecção não destrutiva do conteúdo de aminoácidos nas folhas.

Nosso estudo revela que há relações interessantes entre os AA e bandas que utilizamos para averiguar essas relações, além do visível, a região do SWIR também pode contribuir para estabelecer relações com diferentes aminoácidos. Huang et al. (2021), afirmam que os

comprimentos de onda característicos do nitrogênio do aminoácido foram distribuídos principalmente na região do infravermelho próximo de ondas longas. A utilização de sensores possibilita oportunidades potenciais não apenas a nível laboratorial, mas também usando sistemas de imagem portáteis e uso de satélites para abranger tal tecnologia para o campo, possibilitando avaliar o desempenho da planta a nível de tecidos, plantas individuais ou em culturas inteiras, utilizando sistemas de alta resolução espectral que possibilita avaliar características bioquímicas e fisiológicas de diversidade de plantas (Vergara-Diaz et al., 2020). A partir dessas informações espectrais a relação entre os teores de AA pode ajudar a identificar genótipos com maior eficiência metabólica, qualidade nutricional ou resiliência a estresses, acelerando programas de melhoramento genético.

Há na literatura estudos que demonstram relação entre proteínas que apresentam picos de absorção em 1.460–1.570 nm e 2.000–2.180 nm (Chelladurai & Jayas, 2014; Shenk et al., 2007), justificando as correlações negativas dos AA com as bandas abrangentes da região do SWIR, já que quanto maior o teor menor será a reflectância. Dessa forma a utilização dessas faixas leva à detecção não destrutiva de proteínas a ser conduzida principalmente usando o sistema de infravermelho de ondas curtas (SWIR) com bandas específicas do espectro (X. Li et al., 2023), já mencionadas anteriormente. Ao relacionar características espectrais com grupos funcionais como C-H, N-H e O-H, conteúdos de água, nitrogênio total, aminoácidos livres, cafeína e teanina podem ser detectados indiretamente (Hong et al., 2021; Rébua et al., 2018; Shen et al., 2022). As características espectrais desses grupos funcionais são relacionadas com faixas do espectro visível ao infravermelho próximo, indicando que a combinação dos espectros visível e infravermelho próximo fornece informações úteis possibilitando melhorar a precisão da detecção de compostos específicos, como os AA em amostras (Long et al., 2024).

## **CONCLUSÃO**

A reflectância das folhas de soja é eficiente para separar os genótipos de soja de acordo com o teor de aminoácidos da folha. As bandas espectrais apresentam relações com os aminoácidos. A banda B21 a B28 apresenta maiores relações positivas com ácido glutâmico, arginina, cistina, isoleucina, leucina, metionina, prolina, treonina e valina. A B12 a B17 relacionam positivamente com ácido glutâmico, histidina, metionina, prolina e treonina. B23 e B26 apresentam correlações positivas com arginina, cistina, prolina, tirosina e valina. De forma negativa: B2 a B10 com ácido aspártico, alanina, arginina, metionina, prolina, serina, tirosina e

valina. B15 a B18 com ácido aspártico, alanina, cistina, metionina, prolina e serina e B27 com glicina, histidina, lisina, metionina, prolina, tirosina e valina.

## REFERENCIAS

- Bhering, L. L. (2017). Rbio: A tool for biometric and statistical analysis using the R platform. *Crop Breeding and Applied Biotechnology*, 17(2), 187–190.
- Chelladurai, V., & Jayas, D. S. (2014). Near-infrared imaging and spectroscopy. In *Imaging with Electromagnetic Spectrum: Applications in Food and Agriculture* (pp. 87–127). Springer.
- da Silva Junior, C. A., Nanni, M. R., Shakir, M., Teodoro, P. E., de Oliveira-Júnior, J. F., Cezar, E., de Gois, G., Lima, M., Wojciechowski, J. C., & Shiratsuchi, L. S. (2018a). Soybean varieties discrimination using non-imaging hyperspectral sensor. *Infrared Physics & Technology*, 89, 338–350.
- da Silva Junior, C. A., Nanni, M. R., Shakir, M., Teodoro, P. E., de Oliveira-Júnior, J. F., Cezar, E., de Gois, G., Lima, M., Wojciechowski, J. C., & Shiratsuchi, L. S. (2018b). Soybean varieties discrimination using non-imaging hyperspectral sensor. *Infrared Physics & Technology*, 89, 338–350. [https://doi.org/https://doi.org/10.1016/j.infrared.2018.01.027](https://doi.org/10.1016/j.infrared.2018.01.027)
- Gill, R. J. (1998). New state record: Redgum lerp psyllid, *Glycaspis brimblecombei*. *California Plant Pest and Disease Report*, 17(1/3), 7–8.
- Guo, B., Sun, L., Jiang, S., Ren, H., Sun, R., Wei, Z., Hong, H., Luan, X., Wang, J., & Wang, X. (2022). Soybean genetic resources contributing to sustainable protein production. *Theoretical and Applied Genetics*, 135(11), 4095–4121.
- Hong, T., Yin, J.-Y., Nie, S.-P., & Xie, M.-Y. (2021). Applications of infrared spectroscopy in polysaccharide structural analysis: Progress, challenge and perspective. *Food Chemistry: X*, 12, 100168.
- Huang, H., Hu, X., Tian, J., Jiang, X., Luo, H., & Huang, D. (2021). Rapid detection of the reducing sugar and amino acid nitrogen contents of Daqu based on hyperspectral imaging. *Journal of Food Composition and Analysis*, 101, 103970.
- Kanchana, P., Santha, M. L., & Raja, K. D. (2016). A review on Glycine max (L.) Merr.(soybean). *World J. Pharm. Pharm. Sci*, 5(1), 356–371.
- Kim, W.-S., Chronis, D., Juergens, M., Schroeder, A. C., Hyun, S. W., Jez, J. M., & Krishnan, H. B. (2012). Transgenic soybean plants overexpressing O-acetylserine sulfhydrylase

- accumulate enhanced levels of cysteine and Bowman–Birk protease inhibitor in seeds. *Planta*, 235, 13–23.
- Li, X., Peng, F., Wei, Z., Han, G., & Liu, J. (2023). Non-destructive detection of protein content in mulberry leaves by using hyperspectral imaging. *Frontiers in Plant Science*, 14, 1275004.
- Li, Z., Li, Z., Fairbairn, D., Li, N., Xu, B., Feng, H., & Yang, G. (2019). Multi-LUTs method for canopy nitrogen density estimation in winter wheat by field and UAV hyperspectral. *Computers and Electronics in Agriculture*, 162, 174–182.
- Liang, D., Zhou, Q., Ling, C., Gao, L., Mu, X., & Liao, Z. (2023). Research progress on the application of hyperspectral imaging techniques in tea science. *Journal of Chemometrics*, 37(6), e3481.
- Long, T., Tang, X., Liang, C., Wu, B., Huang, B., Lan, Y., Xu, H., Liu, S., & Long, Y. (2024). Detecting bioactive compound contents in Dancong tea using VNIR-SWIR hyperspectral imaging and KRR model with a refined feature wavelength method. *Food Chemistry*, 460, 140579.
- Mao, Z.-H., Deng, L., Duan, F.-Z., Li, X.-J., & Qiao, D.-Y. (2020). Angle effects of vegetation indices and the influence on prediction of SPAD values in soybean and maize. *International Journal of Applied Earth Observation and Geoinformation*, 93, 102198.
- Pandey, P., & Prabhakar, R. (2016). An analysis of machine learning techniques (J48 & AdaBoost)-for classification. *2016 1st India International Conference on Information Processing (IICIP)*, 1–6. <https://doi.org/10.1109/IICIP.2016.7975394>
- Rébufa, C., Pany, I., & Bombarda, I. (2018). NIR spectroscopy for the quality control of *Moringa oleifera* (Lam.) leaf powders: Prediction of minerals, protein and moisture contents. *Food Chemistry*, 261, 311–321.
- Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, 3(3), 430–439.
- Scott, A. J., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 507–512.
- Shen, S., Hua, J., Zhu, H., Yang, Y., Deng, Y., Li, J., Yuan, H., Wang, J., Zhu, J., & Jiang, Y. (2022). Rapid and real-time detection of moisture in black tea during withering using micro-near-infrared spectroscopy. *Lwt*, 155, 112970.
- Shenk, J. S., Workman Jr, J. J., & Westerhaus, M. O. (2007). Application of NIR spectroscopy to agricultural products. In *Handbook of near-infrared analysis* (pp. 365–404). CRC Press.

- Shu, M., Zhou, L., Chen, H., Wang, X., Meng, L., & Ma, Y. (2022). Estimation of amino acid contents in maize leaves based on hyperspectral imaging. *Frontiers in Plant Science*, 13, 885794.
- Team, R. C. (2013). *R: A language and environment for statistical computing*.
- Teixeira, W. F., Soares, L. H., Fagan, E. B., da Costa Mello, S., Reichardt, K., & Dourado-Neto, D. (2020). Amino acids as stress reducers in soybean plant growth under different water-deficit conditions. *Journal of Plant Growth Regulation*, 39(2), 905–919.
- Vergara-Diaz, O., Vatter, T., Kefauver, S. C., Obata, T., Fernie, A. R., & Araus, J. L. (2020). Assessing durum wheat ear and leaf metabolomes in the field through hyperspectral data. *The Plant Journal*, 102(3), 615–630.