



UNIVERSIDADE
ESTADUAL DE LONDRINA

CLÁUDIA DOMICIANO TORMENA

**APLICAÇÃO DE MÉTODOS QUIMIOMÉTRICOS NA
ANÁLISE DE DADOS QUÍMICOS DE PRODUTOS
AGRÍCOLAS**

Londrina
2022

CLÁUDIA DOMICIANO TORMENA

**APLICAÇÃO DE MÉTODOS QUIMIOMÉTRICOS NA
ANÁLISE DE DADOS QUÍMICOS DE PRODUTOS
AGRÍCOLAS**

Tese de Doutorado apresentado à Universidade Estadual de Londrina - UEL, como requisito principal para a obtenção do título de Doutora em Química.

Orientador: Prof. Dr^a. Ieda Spacino Scarminio

Londrina
2022

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

T684a Tormena, Cláudia Domiciano.
Aplicação de métodos quimiométricos na análise de dados químicos de produtos agrícolas / Cláudia Domiciano Tormena. - Londrina, 2022.
172 f.: il.

Orientador: Ieda Spacino Scarminio.
Tese (Doutorado em Química) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Química, 2022.
Inclui bibliografia.

1. Quimiometria - Tese. 2. Química analítica - Tese. 3. Metabolômica - Tese. 4. Produtos Agrícolas - Tese. I. Scarminio, Ieda Spacino. II. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Química. III. Título.

CDU 54

CLÁUDIA DOMICIANO TORMENA

**APLICAÇÃO DE MÉTODOS QUIMIOMÉTRICOS NA
ANÁLISE DE DADOS QUÍMICOS DE PRODUTOS
AGRÍCOLAS**

Tese de Doutorado apresentado à Universidade Estadual de Londrina - UEL, como requisito principal para a obtenção do título de Doutora em Química.

BANCA EXAMINADORA

Prof. Orientadora Dr^a Ieda Spacino Scarminio
Universidade Estadual de Londrina – UEL

Prof. Dr Paulo Henrique Março
Universidade Tecnológica Federal do Paraná-
UTFPR

Prof. Dr^a Elen Julcileia Romao Sartori Braz
Universidade Estadual de Londrina – UEL

Prof. Dr^a Audrey Alesandra Stingham Garcia
Lonni
Universidade Estadual de Londrina – UEL

Prof. Dr Sérgio Luis Costa Ferreira
Universidade Federal da Bahia – UFBA

Londrina, 06 de Setembro de 2022.

AGRADECIMENTOS

À Professora Dr^a Ieda Spacino Scarminio, minha orientadora e amiga que me acompanhou desde meu trabalho de conclusão de curso, mestrado e agora finalizando o ciclo do doutorado. Não tenho palavras suficientes para dizer o quanto é importante em todos os aspectos da minha vida. Muito obrigada pelos ensinamentos acadêmicos e principalmente os de vida.

Ao Professor Dr. Roy Edward Bruns, amigo, mentor e professor que sempre contribuiu de forma esplêndida para meu crescimento acadêmico e de vida.

À Dr^a Elis Daiane Pauli, minha amiga, co-orientadora e uma pessoa muito especial que entrou na minha vida só para acrescentar tanto profissionalmente quanto pessoalmente. Só tenho a agradecer por todas as conversas, cafés da tarde, desabafos e conselhos.

Ao Dr. Gustavo Galo Marcheafave, meu melhor amigo, que me ajudou desde meu trabalho de conclusão de curso. Novamente me faltam palavras para descrever o quanto é importante na minha vida. Agradeço por todas as conversas, conselhos, desabafos e pela paciência.

À toda minha família. Especialmente aos meus pais, Ivo e Marilda, que sempre almejaram que eu tivesse o melhor estudo possível. Minha tia Lúcia e minha avó, Carmelina (*in memoriam*), que sempre foram um apoio muito importante na minha vida.

Aos meus amigos de infância e faculdade que sempre se fizeram presente durante essa fase da minha vida.

Aos professores do curso de Química da UEL e do Programa de Mestrado e Doutorado em Química que contribuíram para minha graduação e formação acadêmica.

Aos colegas LQCN e LEMA e a todos que, de alguma forma, contribuíram para conclusão desse trabalho.

Aos membros da banca pelo aceite e contribuição ao trabalho.

Por fim, agradeço à Deus.

O “nós” é a modéstia da vida acadêmica.

TORMENA, Cláudia Domiciano. **Aplicação de métodos quimiométricos na análise de dados químicos de produtos agrícolas**. 2022. 171 f. Tese (Doutorado em Química) – Universidade Estadual de Londrina, Londrina, 2022.

RESUMO

Com o passar dos anos a indústria alimentícia passou a necessitar de um amplo controle para maximizar a produção garantindo qualidade, segurança alimentar e padronização dos lotes. Dessa forma, as tecnologias estão se adaptando para atender as necessidades do setor agro-alimentar. Neste trabalho o impacto do dimorfismo sexual e alterações metabólicas na qualidade das folhas de erva-mate (*Ilex paraguariensis*) foram investigados por espectroscopia de infravermelho com transformada de Fourier (FT-IR) combinada com as ferramentas quimiométricas de análise de componentes principais (PCA, do inglês *principal component analysis*) e análise de variância – análise de componentes simultânea (do inglês, *ANOVA-simultaneous component analysis*, ASCA), quando as plantas experimentaram diferentes disponibilidades de luz durante duas fenofases da planta. De fato, o dimorfismo sexual teve uma influência significativa nas impressões digitais espectrais FT-IR da erva-mate apresentando metabólitos como a cafeína para plantas masculinas e feofitina *a* para fêmeas como possíveis biomarcadores. Além disso, foi apontada alteração metabólica nos dois sexos, independente de fatores externos, comprovando que o dimorfismo sexual nas folhas da erva-mate é essencial para aumentar a compreensão do papel metabólico no funcionamento da planta e na qualidade do produto obtido. Este trabalho também apresenta a autenticação de feijão como forma de garantir produtos de maior qualidade para o comércio. Para isso, a análise discriminante por mínimos quadrados parciais (do inglês, *partial least squares discriminant analysis* – PLS-DA) foi aplicada a imagens digitais a fim de desenvolver uma metodologia não destrutiva que permite a discriminação de três cultivares de *Phaseolus vulgaris* L. (Agro ANfc9, IPR-Andorinha e IPR-Sabiá) com diferentes características tecnológicas. A PCA resultou na separação dessas cultivares, mas com certa sobreposição, enquanto a análise supervisionada mostrou que três modelos PLS1-DA, um modelo para cada duas cultivares, foi moderadamente melhor do que o tratamento simultâneo de todas as três cultivares (PLS2-DA). As melhores classificações foram encontradas para as cultivares Agro ANfc9 e IPR-Sabiá do que para IPR-Andorinha. O modelo Agro ANfc9 – IPR-Sabiá classificou corretamente 100% das duas classes de feijão nos conjuntos de treinamento e teste mostrando que essa estratégia analítica é rápida, econômica, ecologicamente correta e pode ser aplicada no controle de qualidade do feijão, ajudando na autenticidade do cultivo para o comércio. Ademais, aspectos da interface entre disponibilidade de luz sobre o estrato vegetal e sensibilidade a biorreguladores em *Coffea arabica* foi investigada usando métodos analíticos clássicos como espectroscopia aliada a quimiometria. Os espectros foram analisados usando uma modificação do método de análise de variância-dimensões comuns, AComDim-ICA (do inglês, *analysis of variance-common dimensions analysis-independent components analysis*). Isso pode facilitar a identificação de variações nos dados resultantes dos efeitos dos fatores de delineamento experimental, biorreguladores e estratificação, bem como suas interações. Por ser um fruto climatérico, o café é sensível à aplicação de etileno exógeno (Ethephon), que interfere no processo de amadurecimento e pode apresentar quantidades elevadas de cafeína no estrato superior da copa (exposto à luz solar intensa). E, a aplicação externa de giberelina resultou no aumento da produção de substâncias da classe fenólica nos grãos de café cru. Com isso pode-se identificar a qualidade dos grãos crus produzidos, a partir do aumento da maturidade das bagas.

Palavras-chave: quimiometria; erva-mate; cultivares de feijões; *Coffea arabica*; PCA; ASCA; PLS-DA; AComDim-ICA.

TORMENA, Cláudia Domiciano. **Application of chemometric methods in agricultural products chemical data**. 2022. 171 p. Tese (Doutorado em Química) – Universidade Estadual de Londrina, Londrina, 2022.

ABSTRACT

Over the years, the food industry began to need control to maximize production, ensuring quality, food safety and batch standardization. In this way, technologies are adapting to meet the needs of the agri-food sector. In this work, the impact of sexual dimorphism and metabolic changes on the quality of yerba mate (*Ilex paraguariensis*) leaves were investigated by Fourier transform infrared spectroscopy (FT-IR) combined with chemometric tools such as the principal component analysis (PCA) and analysis of variance – simultaneous component analysis (ASCA), when plants experienced different light availabilities during two plant phenophases. Indeed, sexual dimorphism had a significant influence on FT-IR spectral fingerprints of yerba-mate presenting metabolites such as caffeine for male plants and pheophytin *a* for females as possible biomarkers. In addition, metabolic changes were identified in both sexes, regardless of external factors, proving that sexual dimorphism in the leaves of yerba-mate is essential to increase the understanding of the metabolic role in the functioning of the plant and in the quality of the final product obtained. This work also presents bean authentication as a way of resulting in higher quality products for the trade. For this, partial least squares discriminant analysis (PLS-DA) was applied to digital images in order to develop a methodology that allows non-destructive discrimination of three cultivars of *Phaseolus vulgaris* L. (Agro Anfc9, IPR-Andorinha and IPR-Sabiá) with different technological characteristics. PCA resulted in the separation of these cultivars, but with some overlap while supervised analysis showed that three models PLS1-DA, one model for two cultivars, was moderately better than the simultaneous treatment of all three cultivars (PLS2-DA). The best classifications were found for the cultivars Agro Anfc9 and IPR-Sabiá than for IPR-Andorinha. The Agro Anfc9 – IPR-Sabiá model correctly classified 100% of the two bean classes in the training and test sets, showing that this analytical strategy is fast, economical, ecologically correct and can be applied in the quality control of beans, helping in the authenticity of the bean cultivation for trade. Furthermore, aspects of the interface between light availability on the plant stratum and sensitivity to bioregulators in *Coffea arabica* were investigated using classical analytical methods such as spectroscopy and chemometrics. The spectra were analyzed using a modification of the analysis of variance-common dimensions method, AComDim-ICA. This can make it easier to identify in the data the effects of experimental design, bioregulator and stratification factors, as well as their interactions. As it is a climacteric fruit, coffee is sensitive to the application of exogenous ethylene (Ethephon), which interferes with the ripening process and may present high amounts of caffeine in the upper stratum of the crown (exposed to intense sunlight). And, external application of gibberellin resulted in increased production of substances of the phenolic class in raw coffee beans. With this, it is possible to identify the quality of the raw grains produced, from the increase in the maturity of the berries.

Key words: chemometrics. yerba-mate; bean cultivars; *Coffea arabica*; PCA; ASCA; PLS-DA; AComDim-ICA.

LISTA DE FIGURAS

CAPÍTULO I

- Figura 1** – Esquema do processo de obtenção de amostras foliares de dois gêneros de erva-mate (*Ilex paraguariensis*) cultivadas em dois ambientes de luz diferentes e colhidas em duas fenofases (estações). 39
- Figura 2** – Gráfico dos escores e *loadings* de PC1 e PC2 para as impressões digitais FT-IR do planejamento estatístico de mistura de folhas de erva-mate masculinas (●) e femininas (▼) colhidas na primavera de plantas cultivadas em clareiras e em áreas sombreadas. 45
- Figura 3** – Gráfico dos escores e *loadings* da PC1 e PC2 para o planejamento estatístico de mistura FT-IR – impressões digitais de folhas de erva-mate masculina (●) e feminina (▼) colhidas no inverno de plantas cultivadas em clareiras e em áreas sombreadas. 47
- Figura 4** – Análise ASCA das impressões digitais FT-IR de folhas de erva-mate machos (●) e fêmeas (▼). A) Gráfico dos escores e B) gráfico de *loading* para o efeito do dimorfismo sexual. 48
- Figura 5** – Média do planejamento estatístico de mistura FT-IR – impressões digitais entre 1625 e 1790 cm^{-1} para folhas masculinas e femininas cultivadas em ambientes claros e sombreados e colhidas no inverno. 50
- Figura 6** – Análise ASCA de impressões digitais FT-IR de erva-mate. A) Gráfico dos escores da PC1 do efeito de acessibilidade à luz; B) gráfico dos *loading* do efeito de acessibilidade à luz; C) gráfico dos escores da PC1 do efeito da época de colheita e D) gráfico dos *loading* do efeito da época de colheita. 52
- Figura 7** – Média dos espectros infravermelhos do planejamento estatístico de mistura de folhas de erva-mate de plantas machos e fêmeas cultivadas em dois ambientes de luz e colhidas em duas fenofases. 53

CAPÍTULO II

- Figura 1** – Esquema de aquisição de imagens digitais: (1) Webcam; (2) Fonte de luz LED; (3) Compartimento da amostra; (4) Suporte; (5) Computador para aquisição dos dados. 74
- Figura 2** – Região de interesse (ROI) utilizada a partir de imagens digitais

	adquiridas para as três classes de cultivares de feijão: Agro ANfc9, IPR-Andorinha e IPR-Sabiá.	75
Figura 3	– Histograma para o espaço de cores na escala de cinza, RGB e HSI para todas as amostras dos três cultivares de feijão e suas médias. A) Agro ANfc9; B) Perfil médio do Agro ANfc9; C) IPR-Andorinha; D) Perfil médio do IPR-Andorinha; E) IPR-Sabiá; e F) Perfil médio do IPR-Sabiá.....	79
Figura 4	– Gráfico dos escores da PCA obtidos a partir dos histogramas RGB, tons de cinza e HSI para todas as amostras de feijão: Agro ANfc9 (♦), IPR-Andorinha (★) e IPR-Sabiá (■).	80
Figura 5	– Resíduos Q <i>versus</i> Hotelling T ² para o modelo PLS-DA de classificação de classes de cultivares de feijão. A) Todas as classes; B) Agro ANfc9 <i>versus</i> IPR-Andorinha; C) Agro ANfc9 <i>versus</i> IPR-Sabiá; D) IPR-Sabiá <i>versus</i> IPR-Andorinha. Predição Agro ANfc9 (♦), validação Agro ANfc9 (▲), predição IPR-Andorinha (★), validação IPR-Andorinha (*), predição IPR-Sabiá (■) e validação IPR-Sabiá (▼).	82
Figura 6	– Valores de Y preditos para o modelo PLS2-DA para a classificação das classes dos cultivares Agro ANfc9, IPR-Andorinha e IPR-Sabiá. Predição Agro ANfc9 (♦), validação Agro ANfc9 (▲), predição IPR-Andorinha (★), validação IPR-Andorinha (*), predição IPR-Sabiá (■) e validação IPR-Sabiá (▼).	83
Figura 7	– Valores de y previstos para os modelos binários PLS1-DA: Agro ANfc9 – IPR-Andorinha (A) Agro ANfc9 classe e (B) IPR-Andorinha classe; Agro ANfc9 – IPR-Sabiá (C) classe Agro ANfc9 e (D) classe IPR-Sabiá; e IPR-Andorinha – IPR-Sabiá (E) classe IPR-Andorinha e (F) classe IPR-Sabiá. Predição Agro ANfc9 (♦), validação Agro ANfc9 (▲), predição IPR-Andorinha (★), validação IPR-Andorinha (*), predição IPR-Sabiá (■) e validação IPR-Sabiá (▼).	87

CAPÍTULO III

Figura 1 – Gráficos dos escores na primeira e segunda componentes principais (PCs) das impressões digitais de transmitância NIR de grãos de *Coffea arabica* adquiridos a partir dos extratos do planejamento estatístico de

- mistura. Os grãos foram provenientes de plantas tratadas com água destilada - Controle (▼), Ethephon (◆) ou GA₃ (■).....113
- Figura 2** – A) Gráficos dos escores da primeira e segunda componentes principais (PCs) das impressões digitais UV-Vis de grãos de *Coffea arabica* adquiridos a partir dos extratos do planejamento estatístico de mistura. B) A aproximação dos gráficos dos escores mostra mais claramente as amostras ED. Os feijões eram de plantas tratadas com água destilada - Controle (▼), Ethephon (◆) ou GA₃ (■).....114
- Figura 3** – Saliências dos blocos decompostos ANOVA, para as Componentes Comuns (CC) 1 a 9, em que BIO representa o fator biorregulador, STR o fator de estratificação, BIO×STR a interação entre biorreguladores e estratificação e, finalmente, R a matriz residual. As barras foram E1 (etanol/905 a 1615 nm - vermelho), E2 (etanol/1652 a 1698 nm - azul), ED1 (etanol e diclorometano/905 a 1615 nm - verde), ED2 (etanol e diclorometano/1652 a 1698 nm - roxo).....116
- Figura 4** – Valores F baseados nos Componentes Comuns (CC) 1 a 9, com altas saliências para resíduos calculados a partir do AComDim-ICA de impressões digitais espectrais NIR para grãos de café originados de plantas controle ou plantas tratadas com biorreguladores e colhidas em diferentes estratos ao longo do perfil vertical da planta. BIO representa o fator biorregulador, STR o fator de estratificação, BIO×STR a interação entre biorreguladores e estratificação e R a matriz residual.....117
- Figura 5** – Gráficos dos escores AComDim-ICA das impressões digitais espectrais NIR de grãos de café originados de plantas tratadas com biorreguladores e colhidas nos três estratos verticais de plantas: A) Gráfico dos escores CC1×CC3 para fator biorregulador - (1) Ethephon, (2) controle e (3) GA₃; B) Gráfico dos escores CC1×CC4 para fator de estratificação - (1) o estrato mais baixo - 1 m, (2) o estrato médio - 2 m e (3) o estrato mais alto - 3 m; C) Gráfico dos escores CC1×CC5 para a interação “biorregulador×estratificação” - (1) Ethephon-1 m, (2) Ethephon-2 m, (3) Ethephon-3 m, (4) controle-1 m, (5) controle-2 m, (6) controle-3 m, (7) GA₃-1 m, (8) GA₃-2 m e (9) GA₃-3 m119
- Figura 6** – Gráficos dos *loadings* AComDim-ICA das impressões digitais espectrais NIR de grãos de café originados de plantas tratadas com biorreguladores

e colhidas nos três estratos verticais: A) Gráfico dos *loadings* CC1×CC3 para fator biorregulador, B) Gráfico dos *loadings* CC1×CC4 para fator de estratificação, C) Gráfico dos *loadings* CC1×CC5 para a interação “biorregulador×estratificação”121

Figura 7 – Saliências dos blocos decompostos ANOVA, para Componentes Comuns (CC) 1 a 9 (ver Tabela 1) sendo BIO o fator biorregulador, STR o fator de estratificação, BIO×STR a interação entre biorreguladores e estratificação e, finalmente, R a matriz residual. As barras indicam E (etanol - vermelho) e ED (etanol e diclorometano - azul).122

Figura 8 – Valores F baseados nos Componentes Comuns (CC) com altas saliências para resíduos calculados a partir do AComDim-ICA de impressões digitais metabolômicas espectrais UV-Vis para grãos de café de plantas tratadas com biorreguladores e colhidas dos três estratos verticais. Sendo, BIO o fator biorregulador, STR o fator de estratificação, BIO×STR a interação entre biorreguladores e estratificação e R a matriz residual.123

Figura 9 – Gráficos dos escores AComDim-ICA das impressões digitais metabolômicas espectrais UV-Vis de grãos de café de plantas tratadas com biorreguladores e colhidas nos três estratos verticais: A) Gráfico dos escores CC1×CC6 para fator de estratificação. (1) o estrato mais baixo - 1 m, (2) o estrato médio - 2 m e (3) o estrato mais alto - 3 m; B) Gráfico dos escores CC1×CC5 para interação “biorregulador×estratificação” - (1) Ethephon-1 m, (2) Ethephon-2 m, (3) Ethephon-3 m, (4) controle-1 m, (5) controle-2 m, (6) controle-3 m, (7) GA₃-1 m, (8) GA₃-2 m e (9) GA₃-3 m125

Figura 10 – Gráficos dos *loadings* AComDim-ICA das impressões digitais metabolômicas espectrais UV-Vis de grãos de café originados de plantas tratadas com biorreguladores e colhidos nos três estratos verticais de plantas: A) CC1×CC6 gráfico dos *loadings* para fator de estratificação, B) Gráfico dos *loadings* CC1×CC5 para interação “biorregulador×estratificação”126

LISTA DE TABELAS

CAPÍTULO I

- Tabela 1** – Composições do planejamento estatístico de mistura para quatro componentes: solventes etanol (**e**), acetato de etila (**a**), diclorometano (**d**) e hexano (**h**)40
- Tabela 2** – Planejamento fatorial (2^3) para a investigação dos efeitos do dimorfismo sexual, acessibilidade à luz e estação de colheita na impressão digital FT-IR de folhas de erva-mate.....42
- Tabela 3** – Resultados da análise ASCA com efeito percentual e valores de p para efeitos principais e fator de interação para folhas de erva-mate.....48
- Tabela 4** – Grupos funcionais e modo de vibração FT-IR dos *loadings* das análises PCA e ASCA: impressões digitais do planejamento estatístico de mistura de folhas de erva-mate.53

CAPÍTULO II

- Tabela 1** – Parâmetros de classificação e predição de qualidade para os feijões dos cultivares Agro ANfc9, IPR-Andorinha e IPR-Sabiá pelos modelos PLS2-DA e PLS1-DA comum.....85
- Tabela 2** – Parâmetros do teste de permutação obtidos para o modelo PLS2-DA.86
- Tabela 3** – Parâmetros do teste de permutação obtidos para o modelo PLS1-DA.88

CAPÍTULO III

- Tabela 1** – Delineamento fatorial com dois níveis e três fatores (3^2) para análise da aplicação de biorreguladores e efeitos da estratificação nas impressões digitais metabolômicas de grãos de *Coffea arabica*.105
- Tabela 2** – Proporção do planejamento estatístico de mistura para os quatro componentes etanol, diclorometano, acetona e hexano107

LISTA DE ABREVIATURAS E SIGLAS

a	Acetato de etila
A	Acetona
AComDim	Análise de variância – análise de dimensão comum, do inglês, analysis of variance-common dimensions analysis
AG	Ácido Giberélicos
ANOVA	Análise de variância
ASCA	Análise de componentes simultânea, do inglês, ANOVA-simultaneous component analysis
ATR	Reflectância total atenuada, do inglês, attenuated total reflectance
BIO	Bioregulador
CC	Componentes comuns
CCSWA	Componentes comuns e pesos específicos, do inglês, common components and specific weights
CO ₂	Dióxido de carbono
CVCE	Erro de classificação de validação cruzada, do inglês, cross-validation classification error
d/D	Diclorometano
e/E	Etanol
FT-IR	Infravermelho com transformada de Fourier
h/H	Hexano
HCA	Análise hierárquica, do inglês, hierarchical cluster analysis
HPLC-PDA	Cromatografia líquida de alta eficiência acoplada a detector com arranjos de fotodiodos, do inglês, High Performance Liquid Chromatography with Photodiode-Array Detection
HSI	Matiz-Saturação-Intensidade do inglês, Hue-Saturation-Intensity
ICA	Análise de componentes independentes, do inglês, independent components analysis
k-NN	k-vizinhos mais próximos, do inglês, k-nearest neighbour
LDA	Análise de discriminante linear, do inglês, linear discriminant analysis
LV	Variáveis latentes, do inglês, Latent Variable

MSC	Correção de dispersão multiplicativa, do inglês, Multiplicative Scatter Correction
NIR	Infravermelho próximo, do inglês, near infrared
PC	Componentes principais, do inglês, principal componente
PCA	Análise de componentes principais, do inglês, principal component analysis
PLS-DA	Análise discriminante por mínimos quadrados parciais, do inglês, partial least squares discriminant analysis
RGB	Vermelho-Verde-Azul, do inglês, Red-Green-Blue
RMSEC	Raiz quadrada do erro quadrático médio de calibração, do inglês, root mean square error of calibration
RMSECV	Raiz quadrada do erro quadrático médio de validação cruzada, do inglês, root mean square error of cross-validation
ROI	Região de interesse, do inglês, region of interest
SAF	Sistema Agroflorestal
SIMCA	Modelagem independente suave por analogia de classe, do inglês, soft independent modelling of class analogies
STR	Estratificação
UV-Vis	Espectroscopia Ultravioleta-Visível

SUMÁRIO

	APRESENTAÇÃO	18
	PRODUÇÃO CIENTÍFICA NO PERÍODO (2018-2022)	20
	INTRODUÇÃO GERAL	23
	OBJETIVOS	31
	OBJETIVOS GERAIS.....	31
	OBJETIVOS ESPECÍFICOS.....	31
	CAPÍTULO I	33
1	INTRODUÇÃO	35
2	MATERIAIS E MÉTODOS	39
2.1	COLETA DE FOLHAS DE ERVA-MATE	39
2.2	PREPARAÇÃO DO EXTRATO	40
2.3	MEDIDAS ESPECTRAIS FT-IR.....	41
2.4	ANÁLISE ESTATÍSTICA E <i>SOFTWARE</i>	41
3	RESULTADOS E DISCUSSÕES	43
4	CONCLUSÃO	56
	REFERÊNCIAS	57
	CAPÍTULO II	68
1	INTRODUÇÃO	70
2	MATERIAIS E MÉTODOS	73
2.1	AMOSTRAS	73
2.2	SISTEMA DE AQUISIÇÃO DE IMAGENS DIGITAIS	73
2.3	ANÁLISE MULTIVARIADA.....	75

2.4	ANÁLISE ESTATÍSTICA E <i>SOFTWARE</i>	77
3	RESULTADOS E DISCUSSÕES	78
3.1	ANÁLISE EXPLORATÓRIA	78
3.2	CLASSIFICAÇÃO.....	81
4	CONCLUSÃO	91
	REFERÊNCIAS	92
	CAPÍTULO III	98
1	INTRODUÇÃO	100
2	MATERIAIS E MÉTODOS	104
2.1	AMOSTRAS	104
2.2	EXTRAÇÃO	106
2.3	UV-VIS	107
2.4	INFRAVERMELHO PRÓXIMO (NIR)	108
2.5	ANÁLISE EXPLORATÓRIA PARA ESCOLHA DO SOLVENTE DE EXTRAÇÃO.....	108
2.6	ANÁLISE ACOMDIM-ICA	109
2.7	ANÁLISE ESTATÍSTICA E <i>SOFTWARE</i>	112
3	RESULTADOS	113
3.1	ANÁLISE EXPLORATÓRIA	113
3.2	ACOMDIM-ICA	115
3.2.1	Análise Multibloco NIR	115
3.2.1	Análise Multibloco UV-Vis	122
4	DISCUSSÃO	127
5	CONCLUSÃO	129
	REFERÊNCIAS	130

CONCLUSÃO GERAL	144
ANEXOS	145
ANEXO A – Estruturas moleculares das principais substâncias químicas discutidas no presente trabalho.....	146
ANEXO B – Figuras em alta qualidade	148

APRESENTAÇÃO

Esta tese de doutorado está dividida em três capítulos que envolvem, na ideia central, a utilização da quimiometria aplicada a produtos agrícolas de importância nacional. Os capítulos são apresentados ao leitor na forma de artigos científicos publicados em revistas científicas internacionais.

O primeiro capítulo dispõe a investigação do efeito do dimorfismo sexual nas alterações metabólicas das folhas de *Ilex paraguariensis* (erva-mate) combinando a espectroscopia de infravermelho por transformada de Fourier (FT-IR) com as ferramentas quimiométricas, como a análise de componentes principais (do inglês, *principal component analysis*, PCA) e ANOVA-análise de componentes simultâneos (do inglês, *ANOVA-simultaneous component analysis* - ASCA), investigando plantas expostas a diversas disponibilidades de luz durante duas fenofases de plantas. Os resultados deste trabalho estão publicados na revista “*Microchemical Journal*” (doi: 10.1016/j.microc.2020.105329).

O segundo capítulo apresenta o uso de imagens digitais aliadas a ferramentas quimiométricas para autenticar cultivares de feijão (*Phaseolus vulgaris* L.) com diferentes características tecnológicas. A aquisição das imagens foi realizada em três diferentes cultivares de feijão e os dados foram posteriormente submetidos à análise multivariada de análise de componentes principais (PCA) e análise discriminante por mínimos quadrados parciais (do inglês, *partial least squares discriminant analysis* - PLS-DA). Os resultados deste capítulo foram publicados na revista “*Food Chemistry*” (10.1016/j.foodchem.2021.130349).

O terceiro capítulo relata aspectos da disponibilidade de luz sobre os estratos da copa/planta e sensibilidade a biorreguladores em *Coffea arabica* para

aumentar a qualidade dos grãos crus produzidos, pelo aumento da maturidade das bagas e, conseqüentemente, a qualidade do produto utilizando combinação de métodos espectroquímicos. Os espectros foram analisados usando uma modificação do método de análise de variância-dimensões comuns, onde a etapa usual de análise de componentes principais dentro do ComDim é substituída por uma análise de componentes independentes. Os resultados deste capítulo foram publicados na revista “*Microchemical Journal*” (10.1016/j.microc.2022.107702).

As referências desta tese estão padronizadas de acordo com a norma de cada revista nos quais os artigos foram publicados.

O Anexo A desta tese apresenta as estruturas moleculares dos metabólitos discutidos durante todo o corpo do texto.

O Anexo B apresenta imagens de alta resolução para uma melhor visualização das figuras.

PRODUÇÃO CIENTÍFICA NO PERÍODO (2018-2022)

- 1) TORMENA, C. D.; RUTLEDGE, D. N.; RAKOCEVIC, M.; BRUNS, R.E.; SCARMINIO, I.S.; MARCHEAFAVE, G. G.; PAULI, E.D. Exogenous application of bioregulators in *Coffea arabica* beans during ripening: Investigation of UV-Visible and NIR mixture design-fingerprints using AComDim-ICA. MICROCHEMICAL JOURNAL, v. 181, p. 107702, 2022.
- 2) SILVA, F.B.; TORMENA, C. D.; PAULI, E.D.; DE ALMEIDA, A.G; BERG, A.B.C; RAKOCEVIC, M.; BRUNS, R.E.; SCARMINIO, I.S.; MARCHEAFAVE, G. G.. Time dependent berry maturation for planting density levels in *Coffea arabica* L. beans: Mixture design-fingerprinting using near-infrared transmittance spectroscopy. JOURNAL OF FOOD COMPOSITION AND ANALYSIS, v. 97, p. 103795, 2021.
- 3) TORMENA, C.D; CAMPOS, R.C.S.; MARCHEAFAVE, G.G.; BRUNS, R.E.; SCARMINIO, I.S.; PAULI, E.D. Authentication of carioca common bean cultivars (*Phaseolus vulgaris* L.) using digital image processing and chemometric tools. FOOD CHEMISTRY, v. 364, p. 130349, 2021.
- 4) MARCHEAFAVE, G. G.; TORMENA, C. D.; TERRILE, A.E.; SALAMANCANETO, C.A.R.; SARTORI, E.R.; RAKOCEVIC, M.; BRUNS, R.E.; SCARMINIO, I.S.; PAULI, E.D. Ecometabolic mixture design-fingerprints from exploratory multi-block data analysis in *Coffea arabica* beans from climate changes: Elevated carbon dioxide and reduced soil water availability. FOOD CHEMISTRY, v. 362, p. 129716, 2021.
- 5) MARCHEAFAVE, G. G.; PAULI, E.D.; TORMENA, C. D.; MATTOS, L.E.; DE ALMEIDA, A.G. ; RAKOCEVIC, M.; BRUNS, R.E. ; SCARMINIO, I.S. Irrigated and CO₂ Level Effects on Metabolism in *Coffea arabica* Beans from Mixture Design - Near Infrared Fingerprints. MICROCHEMICAL JOURNAL, v. 152, p. 104276, 2020.
- 6) MARCHEAFAVE, G.G.; PAULI, E.D.; TORMENA, C.D.; ORTIZ, M.C.V. ; DE ALMEIDA, A.G. ; RAKOCEVIC, M.; BRUNS, R.E. ; SCARMINIO, I.S. Factorial design fingerprint discrimination of *Coffea arabica* beans under elevated carbon dioxide and limited water conditions. TALANTA, v. 209, p. 120591, 2020.
- 7) TORMENA, C. D.; PAULI, E.D.; MARCHEAFAVE, G.G.; SCHEEL, G.L.; RAKOCEVIC, M.; BRUNS, R.E.; SCARMINIO, I.S. FT-IR biomarkers of sexual

- dimorphism in yerba-mate plants: seasonal and light accessibility effects. MICROCHEMICAL JOURNAL, v. 158, p. 105329, 2020.
- 8) MARCHEAFAVE, G.G.; TORMENA, C.D.; MATTOS, L.E.; LIBERATTI, V.R.; FERRARI, A.B.S.; RAKOCEVIC, M.; BRUNS, R.E.; SCARMINIO, I.S; PAULI, E.D. The main effects of elevated CO₂ and soil-water deficiency on ¹H NMR-based metabolic fingerprints of *Coffea arabica* beans by factorial and mixture design. SCIENCE OF THE TOTAL ENVIRONMENT, v. 749, p. 142350, 2020.
- 9) MARCHEAFAVE, G.G.; TORMENA, C.D.; PAULI, E.D.; RAKOCEVIC, M.; BRUNS, R.E.; SCARMINIO, I.S. Experimental mixture design solvent effects on pigment extraction and antioxidant activity from *Coffea arabica* L. leaves. MICROCHEMICAL JOURNAL, v. 146, p. 713-721, 2019.
- 10) TORMENA, C.D.; MARCHEAFAVE, G.G.; RAKOCEVIC, M.; BRUNS, R.E.; SCARMINIO, I.S. Sequential mixture design optimization for divergent metabolite analysis: Enriched carbon dioxide effects on *Coffea arabica* L. leaves and buds. TALANTA, v. 191, p. 382-389, 2019.
- 11) MARCHEAFAVE, G.G.; TORMENA, C.D.; AFONSO, S.; RAKOCEVIC, M.; BRUNS, R.E.; SCARMINIO, I.S. Integrated Chemometric Approach to Optimize Sample Preparation for Detecting Metabolic Changes Provoked by Abiotic Stress in *Coffea arabica* L. Leaf Fingerprints. JOURNAL OF THE BRAZILIAN CHEMICAL SOCIETY, v. 30, p. 2085-2094, 2019.
- 12) TORMENA, C.D.; MARCHEAFAVE, G.G.; PAULI, E.D.; BRUNS, R.E.; SCARMINIO, I.S. Potential biomonitoring of atmospheric carbon dioxide in *Coffea arabica* leaves using near-infrared spectroscopy and partial least squares discriminant analysis. Environmental Science and Pollution Research, v. 26, p. 30356-30364, 2019.

INTRODUÇÃO GERAL

Com o avanço do conhecimento nas áreas alimentícias, química analítica, tecnologia de processamento de alimentos, bem como ciência do consumidor e sensorial tornou-se significativamente maior a necessidade do conhecimento dos sistemas alimentares [1]. Portanto, a necessidade de garantir alimentos que possuem alta qualidade, valor nutricional, segurança alimentar, ausência de alterações e adulterações, genuinidade e tipicidade faz com que se busque ferramentas cada vez mais eficazes para investigar e lidar com a matriz química dos alimentos [2].

Os alimentos são materiais complexos que são comumente estudados por engenheiros, tecnólogos, químicos, físicos, microbiologistas e muitos outros profissionais e suas investigações incluem, mas não se limitam à autenticidade, funcionalidade, bioatividade e segurança alimentar [2]. Sendo assim, informações sobre a composição química de uma planta são cruciais para seu desenvolvimento e usabilidade no setor alimentício e agrônomo [3].

Com a evolução de ferramentas de análise instrumental, como métodos espectroscópicos e espectrométricos, cromatográficos e imagens digitais dentro da metabolômica, proteômica e genômica, melhorou substancialmente nossa capacidade de desvendar e entender a complexidade de matrizes alimentares em termos de composição, microestrutura, modalidade sensorial cruzada e atividade biológica [1]. Contudo, esses métodos anteriormente expostos produzem conjuntos de dados complexos e grandes que podem ser mal validados e interpretados usando métodos estatísticos exclusivamente univariados [1,4]. Como os dados metabólicos são complexos e requerem cálculos mais avançados, a quimiometria é considerada uma maneira rápida e eficiente de analisar uma matriz além da dimensionalidade univariada e desvendar as informações ocultas nesses conjuntos de dados [1,3].

A quimiometria é a área que relaciona a aplicação de métodos estatísticos, matemáticos e lógicos com a química [5–7]. Através dos anos ela cresceu e se tornou um campo altamente interdisciplinar [6,7]. A quimiometria e especificamente a análise multivariada, encontram diversas aplicações no controle de qualidade, determinação quantitativa e qualitativa de parâmetros químicos para avaliar produtos alimentícios [8]. Muitas de suas aplicações, juntamente com medidas analíticas convencionais e inovadoras, foram propostas e realizadas para resolver problemas tecnológicos e legislativos de controle de alimentos. Portanto, a quimiometria é útil para preencher as lacunas em dados multidisciplinares que são necessários para gerar conclusões científicas sólidas e para produzir e agregar conhecimento na ciência de alimentos [2].

Os métodos quimiométricos são usados para separar informações úteis do ruído, descobrir correlações ocultas e identificar uma tendência nos dados, baseados nas medidas multivariadas das amostras [9]. Em linhas gerais, os métodos e princípios multivariados mais comuns para análise de alimentos podem ser apresentados como análise exploratória de dados (método não supervisionado); discriminação e classificação; e regressão e previsão (método supervisionado) [8,10].

Dentre os mais conhecidos como métodos supervisionados podem ser citados a análise de discriminante linear (do inglês, *linear discriminant analysis* – LDA), k-vizinhos mais próximos (do inglês, *k-nearest neighbour* – k-NN), análise discriminante por mínimos quadrados parciais (do inglês, *partial least squares discriminant analysis* – PLS-DA), e modelagem independente suave por analogia de classe (do inglês, *soft independent modelling of class analogies* – SIMCA) [11–13]. Enquanto os métodos não supervisionados mais populares são a análise de componentes principais (do inglês, *principal component analysis* – PCA) e análise hierárquica (do inglês, *hierarchical*

cluster analysis – HCA) [10,14]. Essas abordagens são usadas para análise química de alimentos e de outras áreas e sua escolha depende do problema e do tipo de dados experimentais.

A análise não supervisionada promove a identificação de padrões sem a necessidade de informações iniciais sobre a natureza da amostra. Seu objetivo pleno é identificar agrupamentos naturais entre elas [6,15]. Um dos métodos mais comumente utilizados como análise não supervisionada é a PCA e, além disso, é a ferramenta refinada mais valiosa nas mãos de um quimiometrista [14]. Este método exploratório baseia-se na redução de dimensão e emprega decomposição matricial para este fim. Na prática, busca-se no espaço dimensional a direção na qual a variância dos dados é maior. Esta será a direção da primeira componente principal (PC1), uma combinação linear das variáveis originais [15]. Em seguida, determina-se a direção com a segunda maior variância (PC2), com a restrição de que deve ser ortonormal a PC1 (ou seja, também deve ser um vetor de comprimento unitário). A PC3 terá que ser ortogonal a PC1 e a PC2 e assim por diante. Ao final, as componentes principais formarão um conjunto de bases ortonormais. Uma vez determinadas as componentes principais, os dados podem ser projetados em um subespaço de menor dimensão. Como os primeiros componentes principais contêm a maioria da variância nos dados, esta interpretação simplificada permite a detecção de agrupamentos ou padrões, que seriam difíceis de notar na matriz de dados m -dimensional original – contendo possivelmente centenas ou mesmo milhares de variáveis [15].

Outro método de análise não supervisionado é a Análise de Componentes Comuns e Pesos Específicos (do inglês, *Common Components and Specific Weights*, CCSWA ou ComDim) que é baseado na análise de dados multibloco [16]. Esse algoritmo já foi extensamente estudado e existem algumas variações, dentre elas a AComDim, baseado no mesmo conceito do ANOVA-PCA (do inglês, *ANOVA-principal*

component analysis) que têm sido utilizada para para identificar fatores e interações significativas em um projeto experimental [16,17]. A variação desse método calcula primeiramente uma decomposição da matriz de dados original em diferentes matrizes de efeito correspondentes a cada termo de um modelo ANOVA e então realiza uma análise multivariada de cada matriz de efeito separadamente [18]. Essa análise multivariada pode ser realizada tanto usando uma Análise de Componentes Principais (PCA) ou utilizando uma análise de componentes independentes (do inglês, *Independent Component Analysis*, ICA). Essas matrizes são construídas a partir dos níveis dos diversos fatores estudados [19].

O método AComDim usa uma abordagem de Dimensão Comum (ComDim) na qual diferentes matrizes de dados (ou bloco de dados) são analisadas simultaneamente a fim de se encontrar um espaço comum de representação, com cada matriz tendo um peso específico, chamado saliência ou componentes comuns (CC) [18–20]. A técnica baseia-se na soma ponderada das matrizes de variância-covariância da amostra ou matrizes de associação [18]. Uma vez calculado o primeiro CC (CC1), as matrizes são todas deflacionadas e o processo repetido até que o número necessário de CCs seja extraído [20]. Posteriormente, a significância do bloco é estimada com um teste de Fisher para os quais F_i maior que o valor crítico (F_c) da tabela de Fisher são considerados como estando relacionados a fatores ou interações influentes [19].

Já na análise supervisionada cada amostra possui alguma informação inicial e pré-estabelecida e essa informação é utilizada durante o processamento dos dados, sendo seu objetivo pleno é desenvolver um modelo baseado nas informações contidas nas amostras [6]. A análise discriminante PLS-DA é o método mais popular e eficiente entre os métodos que combinam regressão com uma técnica de classificação [21]. O PLS-DA maximiza a separação entre as classes pré-definidas, em vez de explicar

a variação dentro de um conjunto de dados, ou seja, PLS-DA calcula a probabilidade de cada amostra pertencer a uma determinada classe [22]. É um método de regressão PLS convencional onde a matriz de impressões digitais \mathbf{X} é considerada como uma matriz preditora, e uma matriz de resposta \mathbf{Y} especialmente construída compreende variáveis com valores 'zero' ou 'um' e esses códigos indicam se a amostra é de uma classe ou de outra classe [23]. Diferentemente da PCA, os métodos PLS e PLS-DA buscam uma covariância máxima de \mathbf{X} e \mathbf{Y} ocasionando uma pequena rotação no eixo dos componentes principais, e agora os componentes principais são chamados de variáveis latentes (LV) [22]. Uma vantagem de usar o PLS-DA é que se pode entender quais variáveis carregam as informações de separação de classe, mas uma clara desvantagem do PLS-DA é que ele é altamente influenciado pelo número de classes e distribuição do tamanho da amostra por classe [14]. Este fato torna o uso do PLS-DA apropriado apenas quando existem poucas classes, cada uma com muitas amostras a serem diferenciadas [21]. Outras desvantagens do PLS-DA são: ele requer análise de regressão preliminar, que é sensível a valores discrepantes [22].

Dessa forma, conforme mencionado acima, na indústria alimentícia, uma mesma amostra pode gerar uma quantidade de dados muito grande para caracterizar com alto grau de precisão os produtos alimentícios investigados. Isso requer o uso de ferramentas quimiométricas específicas que determinem de forma muito eficiente as relações entre as tabelas de dados.

REFERÊNCIAS

- [1] C.M. Andre, C. Soukoulis, Food quality assessed by chemometrics, *Foods*. 9 (2020) 2–5. doi:10.3390/foods9070897.
- [2] D. Granato, P. Putnik, D.B. Kovačević, J.S. Santos, V. Calado, R.S. Rocha, A.G. Da Cruz, B. Jarvis, O.Y. Rodionova, A. Pomerantsev, Trends in Chemometrics: Food Authentication, Microbiology, and Effects of Processing, *Compr. Rev. Food Sci. Food Saf.* 17 (2018) 663–677. doi:10.1111/1541-4337.12341.
- [3] K. Kucharska-Ambrożej, J. Karpinska, The application of spectroscopic techniques in combination with chemometrics for detection adulteration of some herbs and spices, *Microchem. J.* 153 (2020) 104278. doi:10.1016/j.microc.2019.104278.
- [4] J.J. Roberts, D. Cozzolino, An Overview on the Application of Chemometrics in Food Science and Technology—An Approach to Quantitative Data Analysis, *Food Anal. Methods*. 9 (2016) 3258–3267. doi:10.1007/s12161-016-0574-7.
- [5] Kenneth R. Beebe, *Chemometrics: A Practical Guide*, Wiley-Interscience, 1998.
- [6] M.M.C. Ferreira, *Quimiometria - Conceitos Métodos e Aplicações*, Editora Unicamp, Campinas, 2015.
- [7] B.D.B. Neto, I.S. Scarminio, R.E. Bruns, 25 anos de Quimiometria no Brasil, *Quim. Nova*. 29 (2006) 1401–1406.
- [8] M. Esteki, J. Simal-Gandara, Z. Shahsavari, S. Zandbaaf, E. Dashtaki, Y. Vander Heyden, A review on the application of chromatographic methods, coupled to chemometrics, for food authentication, *Food Control*. 93 (2018) 165–182. doi:10.1016/j.foodcont.2018.06.015.
- [9] A.M. De Souza, R.J. Poppi, Teaching experiment of chemometrics for exploratory analysis of edible vegetable oils by mid infrared spectroscopy and principal

- component analysis: A tutorial, part I, *Quim. Nova.* 35 (2012) 223–229. doi:10.1590/S0100-40422012000100039.
- [10] J.L. Aleixandre-Tudo, L. Castello-Cogollos, J.L. Aleixandre, R. Aleixandre-Benavent, *Chemometrics in food science and technology: A bibliometric study*, *Chemom. Intell. Lab. Syst.* 222 (2022) 104514. doi:10.1016/j.chemolab.2022.104514.
- [11] P. Geladi, B.R. Kowalski, *Partial least-squares regression: a tutorial*, *Anal. Chim. Acta.* 185 (1986) 1–17.
- [12] S. Wold, *Pattern recognition by means of disjoint principal components models*, *Pattern Recognit.* 8 (1976) 127–139. doi:10.1016/0031-3203(76)90014-5.
- [13] G.R. Flåtén, B. Grung, O.M. Kvalheim, *A method for validation of reference sets in SIMCA modelling*, *Chemom. Intell. Lab. Syst.* 72 (2004) 101–109. doi:10.1016/j.chemolab.2004.03.003.
- [14] A. Rácz, D. Bajusz, K. Héberger, *Chemometrics in Analytical Chemistry*, in: Thomas Engel and Johann Gasteiger (Ed.), *Appl. Chemoinformatics Achiev. Futur. Oppor., First*, Wiley-VCH Verlag GmbH & Co. KGaA, 2018: pp. 471–499. doi:10.1002/9783527806539.ch9.
- [15] R.G. Brereton, J. Jansen, J. Lopes, F. Marini, A. Pomerantsev, O. Rodionova, J.M. Roger, B. Walczak, R. Tauler, *Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools*, *Anal. Bioanal. Chem.* 409 (2017) 5891–5899. doi:10.1007/s00216-017-0517-1.
- [16] C. Rébufa, N. Dupuy, I. Bombarda, *Vibrational Spectroscopy AComDim, a multivariate tool to highlighting impact of agroclimatic factors on Moringa oleifera Lam. leaf 's composition from their FTIR-ATR profiles*, *Vib. Spectrosc.* 116 (2021). doi:10.1016/j.vibspec.2021.103297.

- [17] D.N. Rutledge, *Multiblock Methods in Analytical Chemistry*, 2021. doi:10.30744/brjac.2179-3425.letter-dnrutledge.
- [18] D. Jouan-Rimbaud Bouveresse, R.C. Pinto, L.M. Schmidtke, N. Locquet, D.N. Rutledge, Identification of significant factors by an extension of ANOVA-PCA based on multi-block analysis, *Chemom. Intell. Lab. Syst.* 106 (2011) 173–182. doi:10.1016/j.chemolab.2010.05.005.
- [19] N. Delcourt, C. Rébufa, N. Dupuy, N. Boukhdoud, C. Brunel, J. Abadie, I. Giffard, A.M. Farnet-Da Silva, Infrared spectroscopy as a useful tool to predict land use depending on Mediterranean contrasted climate conditions: A case study on soils from olive-orchards and forests, *Sci. Total Environ.* 686 (2019) 179–190. doi:10.1016/j.scitotenv.2019.05.240.
- [20] D.N. Rutledge, Comparison of Principal Components Analysis, Independent Components Analysis and Common Components Analysis, *J. Anal. Test.* 2 (2018) 235–248. doi:10.1007/s41664-018-0065-5.
- [21] L. Yi, N. Dong, Y. Yun, B. Deng, D. Ren, S. Liu, Y. Liang, Chemometric methods in data processing of mass spectrometry-based metabolomics: A review, *Anal. Chim. Acta.* 914 (2016) 17–34. doi:10.1016/j.aca.2016.02.001.
- [22] I. Marquetti, J.V. Link, A.L.G. Lemes, M.B. dos S. Scholz, P. Valderrama, E. Bona, Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee, *Comput. Electron. Agric.* 121 (2016) 313–319. doi:10.1016/j.compag.2015.12.018.
- [23] F.C.G.B.S. Alves, P. Valderrama, Ultraviolet spectroscopy and supervised pattern recognition methods for authentication of transgenic and non-transgenic soybean oils, *Anal. Methods.* 7 (2015) 9702–9706. doi:10.1039/c5ay02440d.

2 OBJETIVOS

2.1. OBJETIVOS GERAIS

Este trabalho tem como objetivo avaliar a eficiência das ferramentas quimiométricas pertinentes para interpretação de dados químicos de produtos agrícolas visando a melhoria tanto no plantio quanto na qualidade final do produto. Para isso utilizou-se de três produtos agrícolas de importância nacional – a erva-mate, o feijão e o café.

2.2. OBJETIVOS ESPECÍFICOS

O Capítulo I o objetivo foi avaliar os dados de plantas de erva-mate obtidos por espectroscopia no infravermelho médio usando PCA seguido do ASCA para investigar os efeitos ocultos na PCA. E, a partir disso, poder definir possíveis biomarcadores de erva-mate que expressam de forma confiável a variação metabólica do dimorfismo sexual quando as plantas experimentam diferentes disponibilidade de luz em duas fenofases.

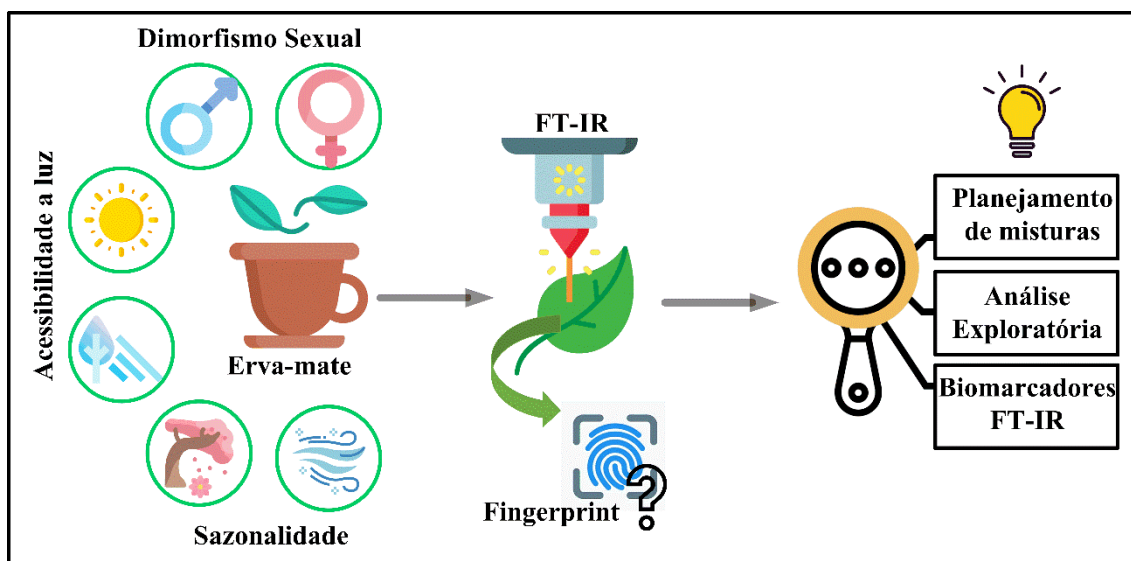
No Capítulo II o objetivo foi avaliar o uso de imagens digitais aliada a métodos quimiométricos supervisionados, como PLS-DA, para autenticação de três cultivares de feijão com diferentes características tecnológicas como, o tempo de cozimento, resistência a doenças, composição química mais desejável e tempo de vida útil, para que assim possa evitar fraudes durante sua comercialização.

Por fim, no Capítulo III o objetivo foi analisar os espectros de infravermelho próximo e ultravioleta de grãos crus de café resultantes da aplicação de biorregulador através de métodos quimiométricos, utilizando a PCA como método exploratório na identificação do melhor solvente extrator e, por fim, empregar uma

modificação do método da análise de variância – análise de dimensão comum a fim de investigar modificações metabólicas refletidas nas impressões espectrais desses grãos de café resultantes da aplicação de um biorregulador no perfil vertical estratificado da planta, e que receberam diferentes quantidades de luz.

CAPÍTULO I

Biomarcadores FT-IR de dimorfismo sexual em plantas de erva-mate: efeitos sazonais e de acessibilidade à luz



RESUMO

Espécies de plantas dióicas, com diferentes gêneros, podem apresentar diferenças no crescimento vegetativo (dimorfismo sexual secundário) e impactar na qualidade de seus produtos. O efeito do dimorfismo sexual nas alterações metabólicas das folhas de *Ilex paraguariensis* (erva-mate) foi investigado por espectroscopia de infravermelho médio com transformada de Fourier (FT-IR) combinada com as ferramentas quimiométricas de análise de componentes principais (PCA) e análise de variância - análise de componentes simultânea (ASCA), quando as plantas experimentaram diferentes disponibilidades de luz durante duas fenofases (estações) da planta. As impressões digitais das folhas da erva-mate foram obtidas por um planejamento estatístico de mistura de quatro componentes (etanol, acetato de etila, diclorometano e hexano). Em folhas de erva-mate colhidas na sombra e em clareiras durante a primavera, a cafeína foi identificada como um possível biomarcador para plantas masculinas, enquanto feofitina *a* foi encontrada para as fêmeas. Folhas colhidas durante o inverno em áreas sombreadas e clareiras mostraram o mesmo padrão de biomarcadores FT-IR da colheita de primavera. O fator de dimorfismo sexual teve uma influência significativa nas impressões digitais espectrais FT-IR da erva-mate confirmando cafeína para plantas masculinas e feofitina *a* para fêmeas como possíveis biomarcadores. Além disso, foi apontada alteração metabólica nos dois sexos, independente de fatores externos. A utilização do planejamento estatístico de mistura foi essencial para a obtenção da máxima diversidade de metabólitos e para garantir maior robustez das análises. O conhecimento dos metabólitos relacionados ao dimorfismo sexual nas folhas da erva-mate é essencial para aumentar a compreensão do papel metabólico no funcionamento da planta e na qualidade do produto final obtido das folhas.

1 INTRODUÇÃO

Biomarcadores são indicadores mensuráveis dos efeitos da exposição que organismos vegetais ou animais sofrem [1]. Quimicamente, eles podem ser considerados como metabólitos que indicam efeitos ambientais e genéticos em organismos vivos [1–4]. A definição do biomarcador acontece quando processos metabólicos sob várias condições são rastreados, modificações genotípicas e fenotípicas são identificadas ou as vias evolutivas de adaptação e a reação de defesa são determinadas [5,6]. Portanto, eles podem ser usados como uma importante ferramenta para monitorar funções biológicas, bioquímicas, fisiológicas e histológicas. Atualmente, no âmbito das adaptações ambientais, são utilizados para melhor compreender a exposição dos organismos vivos aos poluentes ambientais, estresse oxidativo, entre outros [1,7,8].

Recentemente, espectroscopia ultravioleta-visível (UV-Vis) e cromatografia líquida de alta eficiência acoplada a detector com arranjos de fotodiodos (HPLC-PDA) combinadas com ferramentas quimiométricas foram usadas para diferenciar o dimorfismo sexual em erva-mate, identificando as principais classes metabólicas em folhas machos e fêmeas [9]. Essa metodologia mostra que os principais metabólitos marcadores na diferenciação sexual são feofitina α , cafeína, teobromina e derivados dos ácidos clorogênicos. Porém, como já se sabe, HPLC-PDA é uma técnica cara que exige muitos solventes para a fase móvel, alta capacitação e preparação complicada de amostras [10]. Em oposição, a espectroscopia de infravermelho com transformada de Fourier (FT-IR) é uma técnica não invasiva, não destrutiva, rápida e relativamente barata, que fornece informações detalhadas e uma impressão digital geral das amostras [11–13]. Os avanços nos *softwares* de processamento de dados promoveram a aplicação de técnicas de infravermelho na análise de matrizes biológicas complexas. A metodologia FT-IR associada a métodos quimiométricos pode fornecer informações

específicas de forma rápida, direta e confiável a partir de diferentes componentes de amostras complexas [12–15]. Na espectroscopia vibracional, o conjunto de números de ondas associados às ligações químicas e biomoleculares geralmente indica alterações bioquímicas em uma classe de amostra, o que pode ser considerado como a presença de um biomarcador [16].

A análise metabolômica pode resultar em informações sobre modificações químicas em plantas de diferentes gêneros quando submetidas à interação entre estresses abióticos e expressão do dimorfismo sexual [3,17,18]. Ela consiste na amostragem, preparação da amostra, uso de técnicas analíticas instrumentais, processamento e interpretação dos dados. Durante o preparo de amostras complexas, é indicado o uso do delineamento estatístico de misturas, uma vez que a variação das propriedades inerentes aos solventes aumenta a gama de modificações nas diferentes classes dos metabólitos extraídos [19,20]. Na metabolômica, os dados experimentais são obtidos por meio de uma grande quantidade de materiais vegetais, geralmente provenientes de diferentes condições genéticas ou de tratamento [21]. Além disso, o uso de diferentes solventes no processo de extração pode render diferentes resultados, características e perfis químicos sobre um conjunto de metabólitos na matriz em análise [22,23]. Normalmente, a ferramenta estatística clássica usada para dados multivariados é a análise de componentes principais (PCA), que revela fontes de variação dentro do conjunto de dados [21,24,25]. No entanto, a grande variabilidade das condições experimentais implica que os efeitos podem estar ocultos nos resultados da PCA [21,26]. Para investigar esses efeitos, a ANOVA-análise de componentes simultâneos (ASCA) pode ser usada, como uma ferramenta estatística eficaz, para explicar as fontes variacionais nos dados do projeto experimental [21,27,28]. Nessa técnica, a matriz de dados é expressa como uma soma de matrizes de efeitos, de acordo com um planejamento

fatorial. Em seguida, a PCA é realizada nas matrizes de efeitos de interesse, para verificar as fontes de variação para cada efeito específico [21,26,29,30].

A erva-mate (*Ilex paraguariensis* St. Hil.) é uma espécie arbórea dioica com crescimento rítmico [31]. As folhas são utilizadas para o preparo de diversos tipos de bebidas, amplamente consumidas nos países do sul da América Latina [32,33]. A literatura relata a presença de ácidos gálico, caféico, siríngico, ferúlico e *p*-cumárico, rutina, quercetina, canferol e outros compostos bioativos menores nas folhas da erva-mate [34,35]. Porém, seu consumo generalizado se deve aos seus efeitos estimulantes e antioxidantes, principalmente relacionados à presença de compostos polifenólicos (ácidos clorogênicos) e xantinas (cafeína e teobromina) [36]. O dimorfismo sexual secundário (ou seja, caracteres que diferem de gênero que não estão relacionados aos órgãos reprodutivos) desta espécie é expresso na arquitetura global da planta, folha e fisiologia da planta, fenologia [37,38] e níveis de composição química [9]. A diferenciação sexual é óbvia durante os estágios reprodutivos, quando flores e frutos estão presentes nas plantas, enquanto o dimorfismo sexual secundário é menos pronunciado quando as plantas de erva-mate são cultivadas sob a sombra da floresta do que na monocultura [31]. Devido a esse tipo de resposta, a detecção química de alterações metabólicas entre plantas masculinas e femininas pode ser mais precisa. Além disso, a composição química da folha e as características organolépticas da bebida também variam de acordo com a época da colheita e a disponibilidade de luz [39,40].

O estudo da variação do metaboloma nas plantas pode fornecer informações sobre o estado do organismo e os eventos moleculares desencadeados por variações nas condições ambientais e expressões genéticas. Portanto, o objetivo deste estudo foi analisar conjuntos de dados metabolômicos FT-IR de plantas de erva-mate usando métodos quimiométricos, a fim de definir metabólitos biomarcadores que

capturariam de forma confiável a variância metabólica devido ao dimorfismo sexual, quando as plantas experimentaram diferentes disponibilidades de luz durante fenofases das plantas.

2 MATERIAIS E MÉTODOS

2.1. COLETA DE FOLHAS DE ERVA-MATE

Árvores de *Ilex paraguariensis* foram cultivadas na indústria 'Baldo', São Mateus do Sul (25° 52'26"S; 50° 22'58'O), Paraná, Brasil, em um sistema agroflorestal (SAF). As mudas clonais de erva-mate (uma fêmea - F5 e um macho - M9) foram plantadas no sistema agroflorestal em 2005. Dois ambientes foram definidos por 70% de sombreamento e por clareiras, que permitiam mais luz solar direta durante o dia.

As folhas das árvores clonais masculinas e femininas foram colhidas na primavera de 2009, no início do período de rebrota da erva-mate (primeira colheita) e no inverno de 2010, durante a pausa do crescimento (segunda colheita). As folhas foram secas na indústria Baldo, submetidas a uma corrente de ar produzida por ventiladores na sombra a 36 °C, para evitar a decomposição das proteínas. Em seguida, as folhas foram pulverizadas por trituração. Devido à secagem acelerada, a massa foliar obtida manteve o aspecto esverdeado. O esquema de crescimento da planta, coleta de amostra e preparação é ilustrado na Figura 1. O sexo das plantas e a disponibilidade de radiação solar (folhas de clareiras e áreas sombreadas) foram diferenciados durante a colheita de folhas.

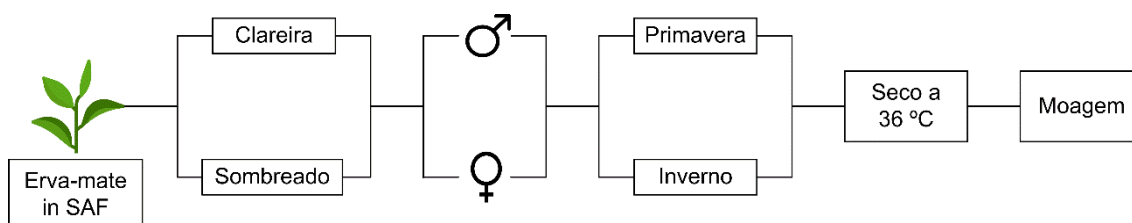


Figura 1. Esquema do processo de obtenção de amostras foliares de dois gêneros de erva-mate (*Ilex paraguariensis*) cultivadas em dois ambientes de luz diferentes e colhidas em duas fenofases (estações).

2.2. PREPARAÇÃO DO EXTRATO

Para a preparação dos extratos, foram utilizados 2,0 g de folhas secas e trituradas, aos quais foram adicionados 6,0 mL da mistura de solventes. As misturas de solventes foram preparadas de acordo com o planejamento de mistura simplex centróide composto por quatro solventes puros, etanol (*e*), acetato de etila (*a*), diclorometano (*d*) e hexano (*h*), todos provenientes de Fmaia, São Paulo, Brasil, (Tabela 1). A mistura quaternária foi realizada em triplicata, totalizando 17 misturas de solventes. A seleção de solventes foi baseada no triângulo de seletividade de solvente de Snyder [41]. O processo de extração foi repetido cinco vezes em banho ultrassônico Unique® Ultra Cleaner TM 1400 (Indaiatuba, Brasil) por 30 minutos, a 25 °C, e filtrado com papel filtro para separar a solução das amostras de folhas de erva-mate, totalizando 30 mL para cada extração. As amostras extraídas foram concentradas e secas sob ventilação de ar corrente.

Tabela 1. Composições do planejamento estatístico de mistura para quatro componentes: solventes etanol (*e*), acetato de etila (*a*), diclorometano (*d*) e hexano (*h*).

Experimento	Notação do extrato	Solventes			
		Etanol	Acetato de etil	Diclorometano	Hexano
1	<i>e</i>	1	0	0	0
2	<i>a</i>	0	1	0	0
3	<i>d</i>	0	0	1	0
4	<i>h</i>	0	0	0	1
5	<i>ea</i>	1/2	1/2	0	0
6	<i>ed</i>	1/2	0	1/2	0
7	<i>eh</i>	1/2	0	0	1/2
8	<i>ad</i>	0	1/2	1/2	0
9	<i>ah</i>	0	1/2	0	1/2
10	<i>dh</i>	0	0	1/2	1/2
11	<i>edh</i>	1/3	0	1/3	1/3
12	<i>eah</i>	1/3	1/3	0	1/3
13	<i>ead</i>	1/3	1/3	1/3	0
14	<i>adh</i>	0	1/3	1/3	1/3
15	<i>eadh - 1</i>	1/4	1/4	1/4	1/4
16	<i>eadh - 2</i>	1/4	1/4	1/4	1/4
17	<i>eadh - 3</i>	1/4	1/4	1/4	1/4

2.3. MEDIDAS ESPECTRAIS FT-IR

Os espectros na região do infravermelho médio (4000 a 675 cm^{-1}) dos extratos brutos foram obtidos com o espectrômetro Thermo Scientific Nicolet iS10 FT-IR utilizando o acessório ATR (refletância total atenuada) com janela Ge. Os espectros foram obtidos com 32 varreduras e resolução de 4 cm^{-1} .

2.4. ANÁLISE ESTATÍSTICA E SOFTWARE

Os espectros brutos foram tratados com pré-processamentos para melhorar a análise dos dados [42]. Antes da aplicação de PCA e ASCA, os dados espectrais FT-IR foram convertidos de transmitância para absorbância ($\log(1/T)$), e a Correção de Dispersão Multiplicativa (MSC) foi aplicada seguida pela centralização média. Para a análise de PCA, foram feitas quatro comparações: 1) sombreado macho×sombreado fêmea na primavera (34×1648); 2) clareira macho×clareira fêmea na primavera (34×1648); 3) sombreado macho×sombreado fêmea no inverno (34×1648) e 4) clareira macho×clareira fêmea no inverno (34×1648). As linhas de cada matriz de comparação correspondem a 34 espectros de extrato do projeto de mistura, 17 para cada gênero (Tabela 1) e as colunas representam os números de onda na faixa de 675-4000 cm^{-1} . Para construir o modelo ASCA uma matriz de dados contendo 136 extratos de erva-mate - 17 impressões digitais para cada gênero em relação a cada combinação do planejamento fatorial experimental (2^3) (dimorfismo sexual, acessibilidade à luz e estação de colheita), codificados de acordo com a Tabela 2, *versus* intensidade espectral (675-4000 cm^{-1}). O tratamento dos dados e a análise quimiométrica foram realizados com o *software* Matlab R2016b[®], usando as ferramentas do PLS toolbox 8.7[®] da Eigenvector Research, enquanto o Origin foi usado para plotar as figuras.

Tabela 2. Planejamento fatorial (2^3) para a investigação dos efeitos do dimorfismo sexual, acessibilidade à luz e estação de colheita na impressão digital FT-IR de folhas de erva-mate.

Fatores	-	+	
1. Dimorfismo sexual	Fêmea	Macho	
2. Acessibilidade à luz	Clareira	Sombreado	
3. Estação de colheita	Primavera	Inverno	
	Nível do fator		
Experimento fatorial	1	2	3
1	-	-	-
2	-	+	-
3	-	-	+
4	-	+	+
5	+	-	-
6	+	+	-
7	+	-	+
8	+	+	+

3. RESULTADOS E DISCUSSÕES

Para definir possíveis biomarcadores para o dimorfismo sexual de dois gêneros de folhas de erva-mate coletadas sob dois microambientes de luz e duas fenofases, foi utilizada a espectroscopia de infravermelho. A utilização de vários solventes através do desenho estatístico de misturas resulta em perfis metabólicos mais informativos, obtendo-se uma maior diversidade de biomarcadores.

O gráfico dos escores da PCA para PC1 e PC2 de folhas de erva-mate cultivadas à sombra e clareira na primavera mostrou separações de impressões digitais FT-IR dos dois gêneros, em ambas as condições de luz (Figura 2). O gráfico dos escores de clareiras para PC1 (69,73%) mostrou que as folhas das plantas femininas tinham valores negativos em sua maioria, enquanto os machos estavam predominantemente no lado positivo (Figura 2A). A PC2 com 13,94% da variância total, não apresentou nenhum padrão de discriminação entre os dois gêneros, possivelmente explicando a variação das impressões digitais devido ao efeito do solvente extrator do desenho da mistura.

O gráfico dos *loadings* da PC1 (Figura 2B) fornece um padrão de discriminação entre as amostras - em uma tentativa de desvendar os biomarcadores metabolômicos nas regiões espectrais FT-IR responsáveis pela discriminação parcial dos escores PCA. *Loadings* mais positivos tiveram maior peso para folhas de plantas masculinas, com bandas correspondentes a metabólitos como cafeína (2920, 2850, 1700, 1658 e 1550 cm^{-1}) [15,43,44] e glicose (2920, 2850, 1065 e 1031 cm^{-1}) [45] (Figura 2B). Os valores dos *loadings* negativos em PC1 (Figura 2B) tiveram maior peso para folhas de plantas femininas, que apresentaram bandas na faixa de 2920 a 2850 cm^{-1} , correspondendo às vibrações de estiramento anti-simétrico e simétrico dos grupos metileno (CH_2) de metabólitos com longas cadeias de CH, como feofitina *a* (1730 cm^{-1}) [46–48].

Na colheita da primavera, os escores da PCA mostraram maior separação entre as impressões digitais masculinas e femininas, quando as plantas da sombra foram comparadas com as das clareiras (Figura 2C). Elas foram discriminadas ao longo da componente principal que explicou 9,53% da variância, PC2, no qual foram encontrados valores positivos para folhas de plantas masculinas e valores negativos para plantas femininas (Figura 2C). Os escores positivos da PC2 tiveram maior peso para folhas de plantas masculinas sombreadas, caracterizadas por bandas como cafeína (2920, 2850, 1700, 1658 e 1550 cm^{-1}) [15,43,44] e glicose (2920, 2850, 1065 e 1031 cm^{-1}) [45], enquanto as cargas negativas representaram bandas com maior peso para o gênero feminino correspondendo a estiramentos de metabólitos como os da feofitina *a* (1730 cm^{-1}) [46–48] e glicose (1065 e 1031 cm^{-1}) [45] (Figura 2D).

As folhas de erva-mate coletadas na sombra e clareiras na primavera mostraram as mesmas regiões de biomarcadores FT-IR (Figura 2B e 2D, respectivamente). Apesar da maior variância total na discriminação das impressões digitais das clareiras (PC1), a formação dos dois grupos correspondentes ao dimorfismo sexual, foi mais evidente para plantas sombreadas (PC2). Esses resultados mostraram que, para plantas sombreadas, todas as impressões digitais do desenho da mistura foram capazes de discriminar os dois gêneros, o que não ocorreu nas clareiras, onde alguns escores não discriminaram a alteração de gênero na primavera.

Primavera

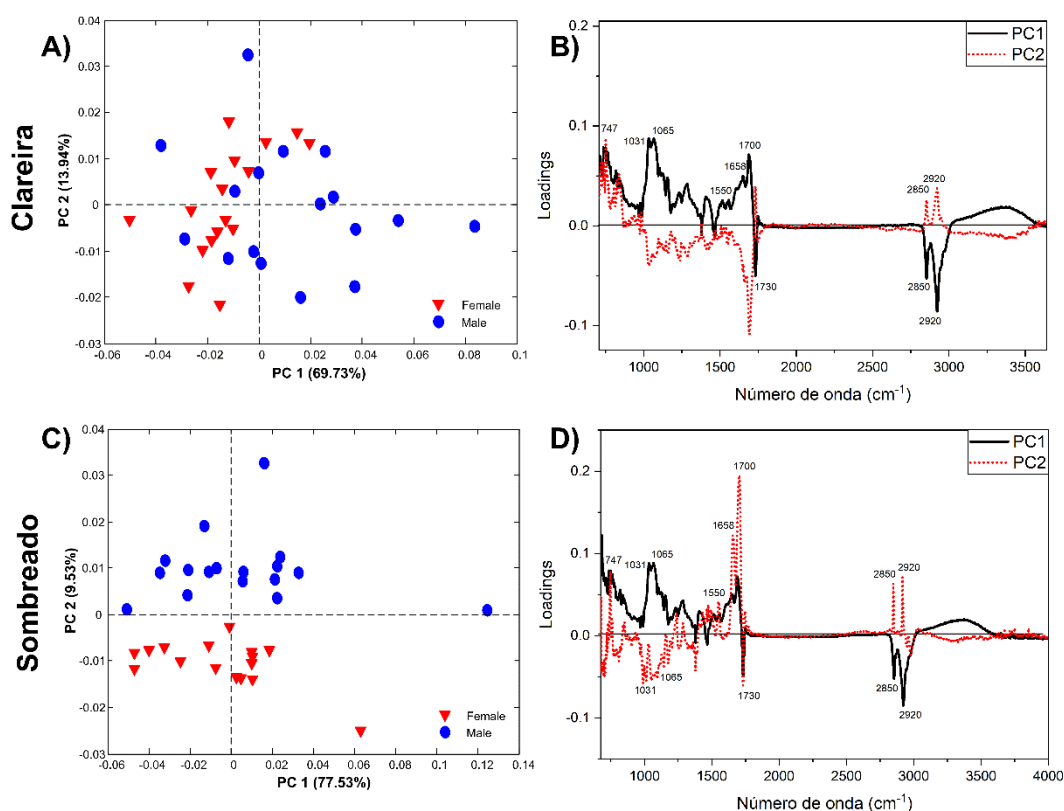


Figura 2. Gráfico dos escores e *loadings* de PC1 e PC2 para as impressões digitais FT-IR do planejamento estatístico de mistura de folhas de erva-mate masculinas (●) e femininas (▼) colhidas na primavera de plantas cultivadas em clareiras e em áreas sombreadas.

A análise das impressões digitais da erva-mate colhidas no inverno em clareiras mostrou discriminação parcial por PC1 (66,26% da variância total) (Figura 3A). A maioria dos valores de escores positivos foram atribuídos às plantas femininas e os valores negativos às masculinas (Figura 3A). Assim, bandas de cafeína (1700, 1658 e 1550 cm⁻¹) [15,43,44] foram mais características para plantas masculinas, assim como bandas de glicose (1065 e 1031 cm⁻¹) [45], enquanto a feofitina *a* (2920, 2850 e 1730 cm⁻¹) [46–48] predomina para plantas femininas (Figura 3B). É possível notar que as faixas mais proeminentes foram as mesmas em plantas colhidas nas clareiras na primavera e no inverno (Figura 2B e 3B).

A PC2 (14,89% da variância total) discriminou a maioria plantas masculinas de plantas femininas, quando estas foram cultivadas à sombra e coletadas no inverno. As plantas masculinas com escores positivos (Figura 3C) correspondem aos números de onda característicos da cafeína (1700, 1658 e 1550 cm^{-1}) [15,43,44], um metabólito chave nesta discriminação (Fig. 3D). Nessa colheita, as folhas das plantas femininas cultivadas à sombra foram discriminadas ao longo do PC2, com escores negativos (Figura 3C), onde as bandas correspondentes à glicose (1065 e 1031 cm^{-1}) [45] foram marcadores (Figura 3D).

A Figura 3 mostra que o padrão de discriminação por PC1 da colheita de plantas no inverno e seu *loadings* são semelhantes aos encontrados para PC2 para plantas coletadas na primavera (Figura 2).

Inverno

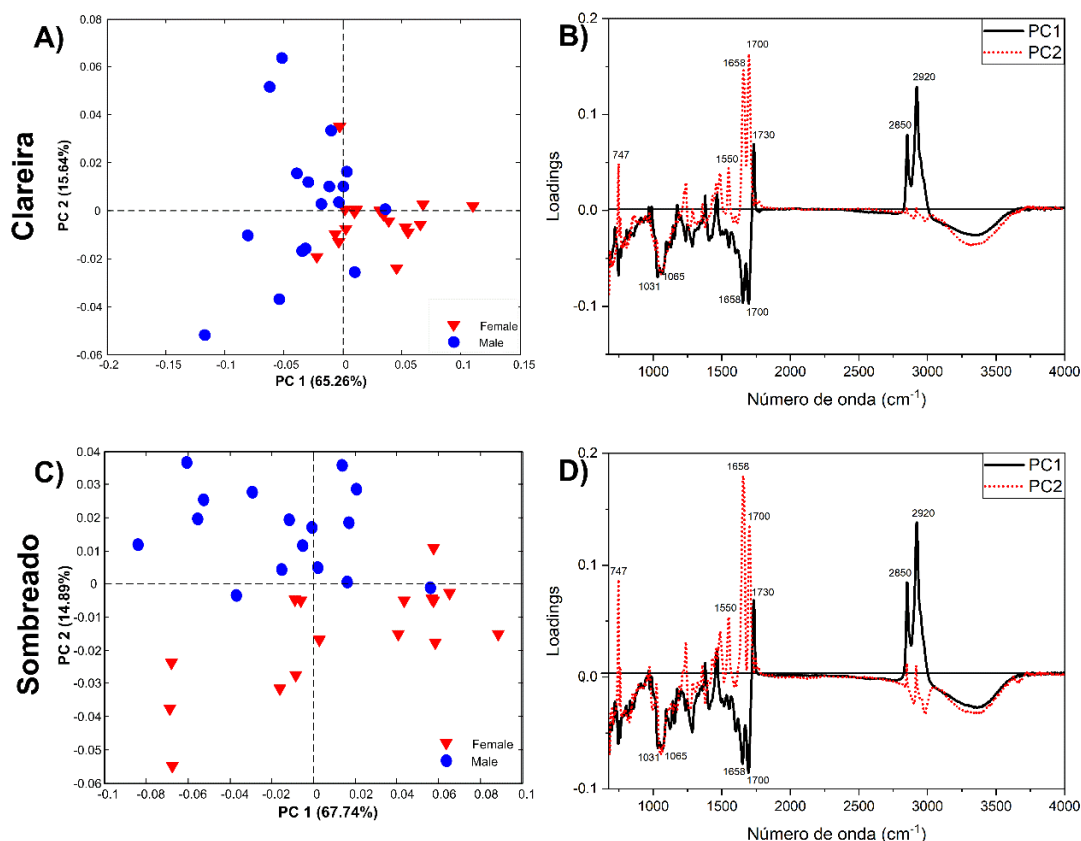


Figura 3. Gráfico dos escores e *loadings* da PC1 e PC2 para o planejamento estatístico de mistura FT-IR – impressões digitais de folhas de erva-mate masculina (●) e feminina (▼) colhidas no inverno de plantas cultivadas em clareiras e em áreas sombreadas.

Para avaliar se o dimorfismo sexual foi de fato um fator estatisticamente significativo na explicação dos fatores do planejamento experimental (dimorfismo sexual, acessibilidade à luz e estação de colheita), ASCA foi aplicada à matriz espectral FT-IR. Os cálculos do valor p foram baseados no teste de permutação e neste estudo foram utilizadas 10.000 randomizações. O fator dimorfismo sexual influenciou significativamente as impressões digitais espectrais FT-IR da erva-mate, com p -valor de 0,0001 (Tabela 3). Não foram encontradas interações binárias e ternárias significativas envolvendo dimorfismo sexual na região FT-IR com os demais fatores, com valores de p maiores que 0,05. Os gráficos dos escores e *loadings* do submodelo PCA para dimorfismo

sexual mostraram bandas em 747; 1658 e 1700 cm^{-1} caracterizando as folhas de plantas machos de erva-mate e as bandas em 1730 cm^{-1} e em torno de 2850 e 2920 cm^{-1} para folhas de plantas femininas (Figura 4).

Tabela 3. Resultados da análise ASCA com efeito percentual e valores de p para efeitos principais e fator de interação para folhas de erva-mate.

	PCs	Efeito (%)	p -valor
Dimorfismo sexual	1	7,39	0,0001
Acessibilidade à luz	1	1,62	0,0307
Estação de colheita	1	28,63	0,0001
Interação (Dimorfismo sexual×Acessibilidade à luz)	1	1,26	> 0,05
Interação (Dimorfismo sexual×Estação de colheita)	1	2,50	> 0,05
Interação (Acessibilidade à luz×Estação de colheita)	1	0,18	> 0,05
Interação (Dimorfismo×Luz×Estação)	1	0,14	> 0,05
Residual		58,28	

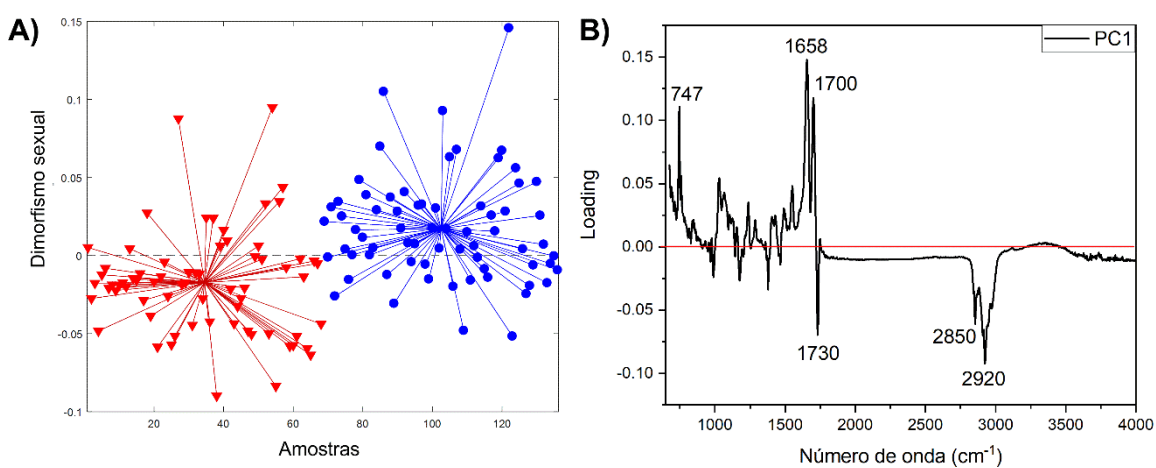


Figura 4. Análise ASCA das impressões digitais FT-IR de folhas de erva-mate machos (●) e fêmeas (▼). A) Gráfico dos escores e B) gráfico de *loading* para o efeito do dimorfismo sexual.

PCA (Figura 2 e Figura 3) e modelos ASCA (Tabela 3 e Figura 4) indicam mudanças nas impressões digitais FT-IR de folhas de erva-mate relacionadas ao dimorfismo sexual, mostrando que as plantas masculinas e femininas atuaram na promoção da biossíntese de alguns metabólitos mais do que outros, dependendo do gênero. Bandas de cafeína (1658 e 1700 cm^{-1}) e de 747 cm^{-1} distinguiram as folhas das plantas masculinas, para todos os métodos quimiométricos aqui aplicados, enquanto a feofitina *a* (1730 cm^{-1}) correspondeu à caracterização das folhas femininas. A Figura 5 mostra que as bandas para as plantas femininas e masculinas sombreadas têm intensidades muito semelhantes em 1730 cm^{-1} , embora o sinal seja mais evidente para as folhas femininas. Evidentemente, isso ocorre devido à sobreposição das bandas de cafeína serem consideravelmente menos intensas em 1700 e 1658 cm^{-1} em plantas femininas quando comparadas com plantas masculinas. A banda da feofitina *a* foi mais proeminente nas folhas das plantas femininas, explicando porque os *loadings* para este metabólito foram maiores para todos os métodos quimiométricos aplicados nesta investigação. Isso pode estar relacionado a atividades fotossintéticas mais altas em plantas femininas em comparação com os machos [31], o que muito provavelmente está relacionado à capacidade de transferência de elétrons do fotossistema II e/ou conteúdo de clorofila em complexos de centros de reação [49].

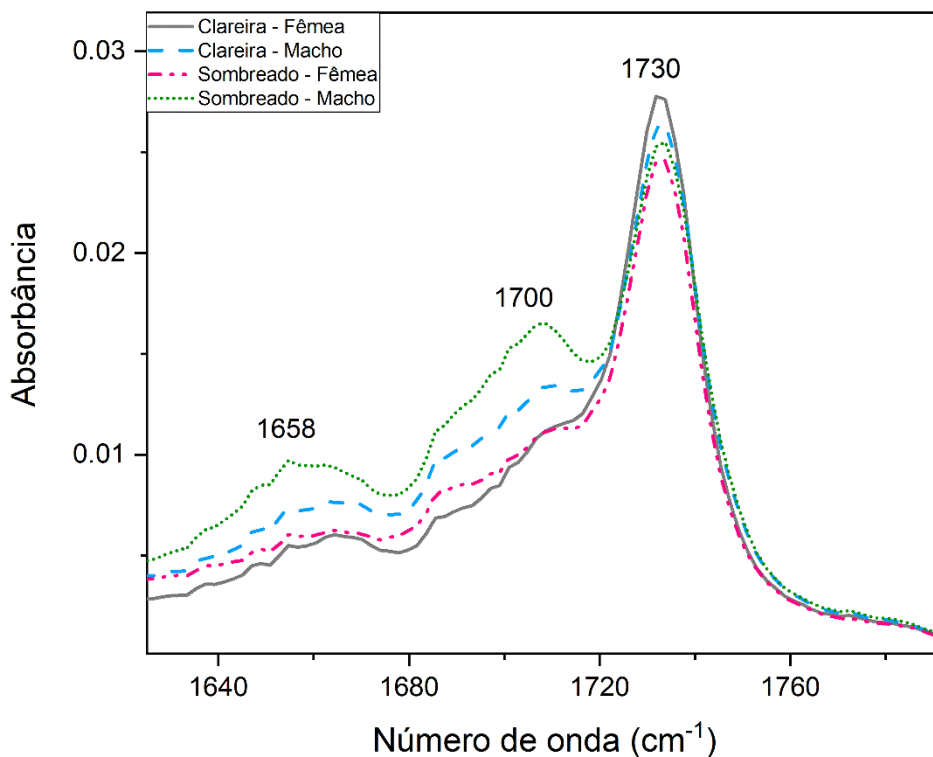


Figura 5. Espectro médio de absorvância FT-IR – impressões digitais entre 1625 e 1790 cm^{-1} para folhas masculinas e femininas cultivadas em ambientes claros e sombreados e colhidas no inverno.

O teste de permutação indicou significância estatística para os fatores de acessibilidade à luz e estação de colheita na impressão digital FT-IR ($p_{\text{acessibilidade à luz}} = 0,0307$ e $p_{\text{estação de colheita}} = 0,0001$). Isso indica que, para cada gênero de erva-mate, mudanças na acessibilidade à luz ou época de colheita provocaram alterações nas impressões digitais metabólicas. O efeito do fator de acessibilidade à luz na mudança do nível das impressões digitais é de 1,62% - menos do que os efeitos do dimorfismo sexual e da época de colheita. Embora significativo, esse efeito menor pode ser parcialmente visto nos escores da primeira componente (Figura 6A). Os gráficos de *loadings* de acessibilidade à luz (Figura 6B) mostraram bandas correspondentes a estiramentos da cafeína (1658 e 1700 cm^{-1}) e em 747 cm^{-1} caracterizando folhas de plantas de erva-mate

sombreadas, enquanto bandas de 1730 cm^{-1} (relacionadas à feofitina *a*) e vibrações de estiramento anti-simétrico e simétrico de grupos metileno (CH_2) (2850 e 2920 cm^{-1}) caracterizaram folhas de plantas de erva-mate em clareiras. De acordo com esses resultados, Coelho et al., [50] e Rossa et al., [51] descobriram que o conteúdo de metilxantina, incluindo cafeína, aumenta com a redução da luz.

Embora esse não seja o objetivo principal deste estudo, a época de colheita apresentou o maior efeito significativo entre os fatores do planejamento fatorial (28,63%). Mudanças no perfil químico espectroscópico da primavera ao inverno ocorreram nas folhas masculinas e femininas. Os escores projetados ajudaram a distinguir as impressões digitais da primavera do inverno (Figura 6C). O gráfico de *loadings* (Figura 6D) mostrou a forte influência das bandas características de cafeína (1658 e 1700 cm^{-1}) para a colheita de inverno e feofitina *a* (1730 cm^{-1}) e vibrações de estiramento anti-simétrico e simétrico de grupos metileno (CH_2) (2850 e 2920 cm^{-1}) para a colheita da primavera. Estudos anteriores mostraram a influência do período de colheita no perfil metabólico da erva-mate [52]. Como a erva-mate tem um crescimento rítmico [31], no fluxo de crescimento da primavera, o teor de cafeína é menor e quando as folhas envelhecem, durante a pausa de crescimento do inverno, o teor de cafeína é maior [52]. Os números de onda e seus respectivos grupos funcionais, modos vibracionais e metabólitos sugeridos estão resumidos na Tabela 4. A média dos espectros na região do infravermelho das impressões digitais do projeto de mistura para oito condições são apresentados na Figura 7.

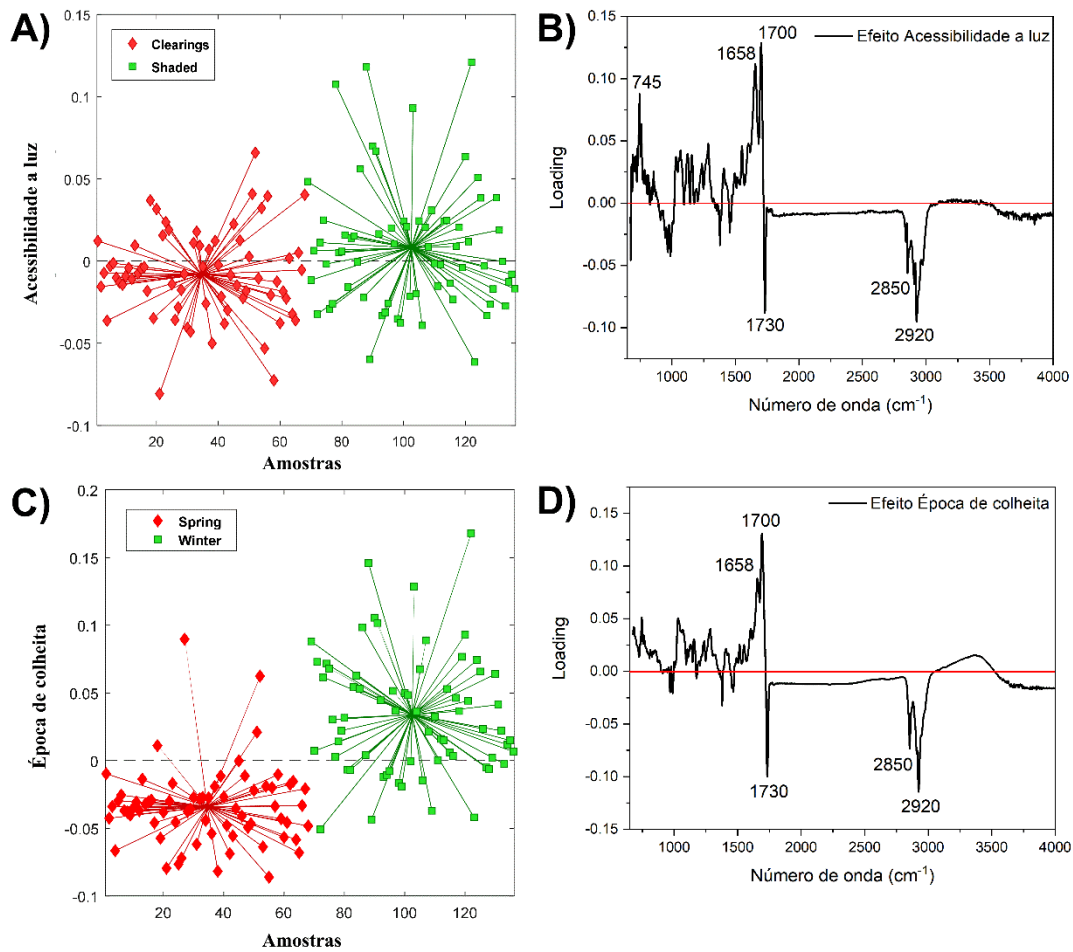


Figura 6. Análise ASCA de impressões digitais FT-IR de erva-mate. A) Gráfico dos escores da PC1 do efeito de acessibilidade à luz; B) gráfico dos *loading* do efeito de acessibilidade à luz; C) gráfico dos escores da PC1 do efeito da época de colheita e D) gráfico dos *loading* do efeito da época de colheita.

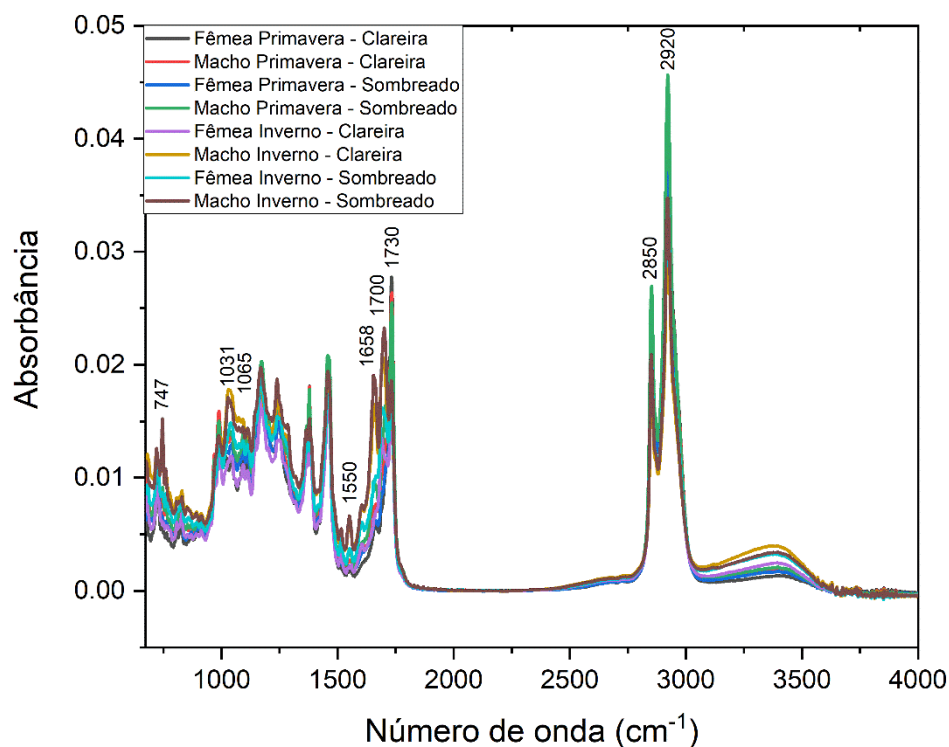


Figura 7. Média dos espectros infravermelhos do planejamento estatístico de mistura de folhas de erva-mate de plantas machos e fêmeas cultivadas em dois ambientes de luz e colhidas em duas fenofases.

Tabela 4. Grupos funcionais e modo de vibração FT-IR dos *loadings* das análises PCA e ASCA: impressões digitais do planejamento estatístico de mistura de folhas de erva-mate.

Número de onda (cm ⁻¹)	Atribuições de ligações	Modo de vibração	Sugestão de metabólito
747	Desconhecido	Desconhecido	Desconhecido
1031	Ligação C-O e C-C	Estiramento	Glicose
1065	Ligação C-O e C-C	Estiramento	Glicose
1550	Ligação C-N	Estiramento	Cafeína
1658	Ligação vibracional C=O, Ligação cíclica C=C de hidrocarbonetos e Ligação C=N	Estiramento	Cafeína
1700	Ligação vibracional C=O, Ligação cíclica C=C de hidrocarbonetos e Ligação C=N	Estiramento	Cafeína
1730	Ligação C=O éster e	Estiramento	Feofitina <i>a</i>

C=O cetona			
2850	Ligação de grupo -CH ₂	Estiramento	Cafeína, Glicose, Feofitina <i>a</i>
2920	Ligação de grupo -CH ₂	Estiramento	Cafeína, Glicose, Feofitina <i>a</i>

As espécies vegetais podem apresentar diferentes metabólitos e/ou proporções em relação ao dimorfismo sexual. Além disso, as plantas podem responder de diferentes maneiras às condições ambientais e estresses abióticos. Para feofitina *a*, um primeiro intermediário portador de elétrons na via de transferência de elétrons do fotossistema II [49] e também um produto de degradação da clorofila devido à substituição de Mg²⁺ por H⁺ [53], nenhuma resposta unânime foi encontrada na literatura em relação a dimorfismo sexual e sua interação com o ambiente de cultivo [38,54,55]. Em relação à erva-mate, as abundâncias de feofitina *a* nas plantas masculinas foram maiores do que nas femininas, independentemente das condições de acesso à luz e diferentes períodos de colheita [9]. Para as plantas masculinas, a cafeína foi importante por sua discriminação em relação às fêmeas. Maior conteúdo de cafeína é encontrado em machos cultivados em áreas abertas do que em fêmeas [4]. Além disso, em relação às plantas masculinas, uma faixa desconhecida em 747 cm⁻¹ caracterizou essas folhas. Estudos que relacionam metabólitos ao dimorfismo sexual em folhas de erva-mate ainda são escassos, e esse conhecimento é essencial para melhor compreender o papel metabólico no funcionamento da planta e na qualidade final do produto obtido das folhas. No entanto, a diferença entre os metabólitos das plantas masculinas e femininas pode ocorrer devido à maior necessidade de as plantas femininas alocarem seus recursos ao processo reprodutivo. Esse requisito pode impactar negativamente a taxa de crescimento feminino e longevidade quando comparado aos machos [56], mas também pode aumentar sua força fisiológica para completar todas as fenofases da planta [31].

Diante desses resultados, o planejamento estatístico de mistura, como

ferramenta analítica na busca por biomarcadores FT-IR, é essencial para atingir o máximo de metabólitos extraídos e, conseqüentemente, impressões digitais, que correspondem ao metaboloma real. A feofitina *a* é mais fácil de extrair em solventes polares, como o etanol. Por outro lado, solventes apolares, como o hexano, tendem a extrair minimamente a feofitina *a*, ao contrário dos lipídios [57,58] - essa analogia é um exemplo de alteração de perfis químicos e, conseqüentemente, a importância de maior domínio experimental na variação das impressões digitais. A combinação de todas as impressões digitais possíveis, combinada com ferramentas quimiométricas, permite inferência global da mudança metabolômica dos sistemas da planta e fornece o máximo de discriminadores da planta com maior precisão. O uso de um único solvente no processo de extração não denota necessariamente dados errôneos, mas não satisfaz um domínio experimental robusto. Em sistemas de extração de metabólitos secundários, isso pode ser otimizado pelo planejamento de misturas, onde as variações nas concentrações das espécies são correlacionadas com a variação das proporções do solvente da mistura [59,60].

4. CONCLUSÃO

Este estudo fornece com sucesso uma abordagem metabolômica empregando impressões digitais de planejamento estatístico de mistura de espectroscopia de infravermelho com transformada de Fourier (FT-IR) para investigar biomarcadores de dimorfismo sexual em folhas de erva-mate, quando as plantas experimentaram várias disponibilidades de luz durante duas fenofases de plantas. As impressões digitais FT-IR juntamente com a análise de componentes principais (PCA) e ANOVA-análise de componentes simultâneos (ASCA) foram úteis para explorar e determinar a significância estatística dos efeitos principais e de interação nas mudanças nas impressões digitais devido à evolução natural das fenofases e variações na disponibilidade de luz em um sistema agroflorestal. A análise exploratória PCA mostrou que os biomarcadores FT-IR de dimorfismo sexual permaneceram constantes, independentemente da fenofase em ambos os níveis de acessibilidade à luz. Em folhas de erva-mate colhidas à sombra e em clareiras, a cafeína foi identificada como possível biomarcador para plantas masculinas, enquanto a feofitina *a* foi identificada nas fêmeas. As bandas dos biomarcadores FT-IR foram 2920, 2850, 1700, 1658 e 1550 cm^{-1} para cafeína e 2920, 2850 e 1730 cm^{-1} para feofitina *a*. Dentre os fatores observados, a época de colheita (fenofase) foi identificada como a maior fonte de variância das impressões digitais FT-IR, com 28,63%, seguida do dimorfismo sexual com 7,39% e acessibilidade à luz com 1,62% de acordo com a análise ASCA. O uso de projetos de mistura estatística pode incluir tantos perfis metabólicos quanto possível, permitindo uma diversidade ainda maior de biomarcadores nas impressões digitais das folhas de erva-mate. Nas caracterizações do dimorfismo sexual da erva-mate, a mistura FT-IR de impressões digitais mostrou-se vantajosa, permitindo resultados altamente satisfatórios quando comparados a técnicas analíticas mais caras.

REFERÊNCIAS

- [1] R.J. Hugget, R.A. Klmerle, P.M. Mehrle, H.L. Bergman, Biomarkers: biological, physiological, and histological markers of anthropogenic stress, 1 st, CRC Press, Boca Raton, FL, 1992.
- [2] C.D. Tormena, G.G. Marcheafave, E.D. Pauli, R.E. Bruns, I.S. Scarminio, Potential biomonitoring of atmospheric carbon dioxide in *Coffea arabica* leaves using near-infrared spectroscopy and partial least squares discriminant analysis, Environ. Sci. Pollut. Res. (2019). doi:10.1007/s11356-019-06163-1.
- [3] M.J. Palumbo, F.E. Putz, S.T. Talcott, Nitrogen fertilizer and gender effects on the secondary metabolism of yaupon, a caffeine-containing north American holly, Oecologia. 151 (2007) 1–9. doi:10.1007/s00442-006-0574-1.
- [4] M. Rakocevic, M. Janssens, R. Schere, Light responses and gender issues in the domestication process of yerba-mate, a subtropical evergreen, in: A.D. Bezerra, T.S. Ferreira (Eds.), Evergreens Types, Ecol. Conserv., UK ed, Nova Science Publishers, Inc, 2012: pp. 63–96.
- [5] I. V Kosakivska, Biomarkers of plants with different types of ecological strategies, Gen. Appl. Plant Physiol. 34 (2008) 113–126.
- [6] Y. Okazaki, K. Saito, Recent advances of metabolomics in plant biotechnology, Plant Biotechnol. Rep. 6 (2012) 1–15. doi:10.1007/s11816-011-0191-2.
- [7] M.E. Yonny, A. Rodríguez Torresi, C. Cuyamendous, G. Réversat, C. Oger, J.M. Galano, T. Durand, C. Vigor, M.A. Nazareno, Thermal Stress in Melon Plants:

- Phytosteranes and Phytosterols as Oxidative Stress Biomarkers and the Effect of Antioxidant Supplementation, *J. Agric. Food Chem.* 64 (2016) 8296–8304. doi:10.1021/acs.jafc.6b03011.
- [8] G. Glauser, D. Guillarme, E. Grata, J. Boccard, A. Thiocone, P.A. Carrupt, J.-L. Veuthey, S. Rudaz, J.-L. Wolfender, Optimized liquid chromatography-mass spectrometry approach for the isolation of minor stress biomarkers in plant extracts and their identification by capillary nuclear magnetic resonance, *J. Chromatogr. A* 1180 (2008) 90–98. doi:10.1016/j.chroma.2007.12.021.
- [9] E.D. Pauli, G.L. Scheel, F. Delarozza, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Photodiode array chromatographic-spectrophotometric metabolite quantification for yerba-mate plant sexual dimorphism differentiation, *Microchem. J.* 151 (2019) 104218. doi:10.1016/j.microc.2019.104218.
- [10] I. Nugrahani, E.Y. Manosa, L. Chintya, FTIR-derivative as a green method for simultaneous content determination of caffeine, paracetamol, and acetosal in a tablet compared to HPLC, *Vib. Spectrosc.* 104 (2019) 102941. doi:10.1016/j.vibspec.2019.102941.
- [11] N. Wang, L.T. Lim, Fourier transform infrared and physicochemical analyses of roasted coffee, *J. Agric. Food Chem.* 60 (2012) 5446–5453. doi:10.1021/jf300348e.
- [12] F.D.S. Grasel, M.F. Ferrão, C.R. Wolf, Development of methodology for identification the nature of the polyphenolic extracts by FTIR associated with

- multivariate analysis, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 153 (2016) 94–101. doi:10.1016/j.saa.2015.08.020.
- [13] M. Bassbasi, M. De Luca, G. Ioele, A. Oussama, G. Ragno, Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data, *J. Food Compos. Anal.* 33 (2014) 210–215. doi:10.1016/j.jfca.2013.11.010.
- [14] J. Moros, S. Garrigues, M. de la Guardia, Vibrational spectroscopy provides a green tool for multi-component analysis, *TrAC - Trends Anal. Chem. Anal. Chem.* 29 (2010) 578–591. doi:10.1016/j.trac.2009.12.012.
- [15] M.M. Paradkar, J. Irudayaraj, A rapid FTIR spectroscopic method for estimation of caffeine in soft drinks and total methylxanthines in tea and coffee, *J. Food Sci.* 67 (2002) 2507–2511. doi:10.1111/j.1365-2621.2002.tb08767.x.
- [16] J. Trevisan, J. Park, P.P. Angelov, A.A. Ahmadzai, K. Gajjar, A.D. Scott, P.L. Carmichael, F.L. Martin, Measuring similarity and improving stability in biomarker identification methods applied to Fourier-transform infrared (FTIR) spectroscopy, *J. Biophotonics.* 7 (2014) 254–265. doi:10.1002/jbio.201300190.
- [17] S. Zhang, S. Lu, X. Xu, H. Korpelainen, C. Li, Changes in antioxidant enzyme activities and isozyme profiles in leaves of male and female *Populus cathayana* infected with *Melampsora larici-populina*, *Tree Physiol.* 30 (2010) 116–128. doi:10.1093/treephys/tpp094.

- [18] M. Juvany, S. Munné-Bosch, Sex-related differences in stress tolerance in dioecious plants: a critical appraisal in a physiological context, *J. Exp. Bot.* 66 (2015) 6083–6092. doi:10.1093/jxb/erv343.
- [19] A.E. Terrile, G.G. Marcheafave, G.S. Oliveira, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Chemometric Analysis of UV Characteristic Profile and Infrared Fingerprint Variations of *Coffea arabica* Green Beans under Different Space Management Treatments, *J. Braz. Chem. Soc.* 27 (2016) 1254–1263. doi:10.5935/0103-5053.20160022.
- [20] C.D. Tormena, G.G. Marcheafave, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Sequential mixture design optimization for divergent metabolite analysis: Enriched carbon dioxide effects on *Coffea arabica* L. leaves and buds, *Talanta*. 191 (2019) 382–389. doi:10.1016/j.talanta.2018.09.002.
- [21] M.E. Timmerman, H.C.J. Hoefsloot, A.K. Smilde, E. Ceulemans, Scaling in ANOVA-simultaneous component analysis, *Metabolomics*. 11 (2015) 1265–1276. doi:10.1007/s11306-015-0785-8.
- [22] P.K. Soares, G.G. Marcheafave, A. de A. Gomes, I.S. Scarminio, R.E. Bruns, Mixture Design PARAFAC HPLC-DAD Metabolomic Fingerprints of Fractionated Organic and Basic Extracts from *Erythrina speciosa* Andrews Leaves, *Chromatographia*. 81 (2018) 1189–1200. doi:10.1007/s10337-018-3554-9.

- [23] P.K. Soares, R.E. Bruns, I.S. Scarminio, Statistical mixture design investigation of fractionated and total extracts from *Erythrina speciosa* Andrews leaves, *J. Sep. Sci.* 32 (2009) 644–652. doi:10.1002/jssc.200800534.
- [24] M.P.H. Verouden, J.A. Westerhuis, M.J. van der Werf, A.K. Smilde, Exploring the analysis of structured metabolomics data, *Chemom. Intell. Lab. Syst.* 98 (2009) 88–96. doi:10.1016/j.chemolab.2009.05.004.
- [25] M.H.F. Spoto, R.E. Domarco, J.M.M. Walder, I.S. Scarmínio, R.E. Bruns, Sensory evaluation of orange juice concentrate as affected by irradiation and storage, *J. Food Process. Preserv.* 21 (1997) 179–191. doi:10.1111/j.1745-4549.1997.tb00775.x.
- [26] W. Cheng, K.M. Sørensen, R.J. Mongi, B.K. Ndabikunze, B.E. Chove, D.W. Sun, S.B. Engelsen, A comparative study of mango solar drying methods by visible and near-infrared spectroscopy coupled with ANOVA-simultaneous component analysis (ASCA), *LWT - Food Sci. Technol.* 112 (2019) 108214. doi:10.1016/j.lwt.2019.05.112.
- [27] N. Cañellas, R. Solà-Alberich, J. Brezmes, R. Mallol, R.-M. Valls, M.A. Rodríguez, M. Vinaixa, A. Anguera, X. Correig, Use of multivariate chemometric algorithms on ¹H NMR data to assess a soluble fiber (*Plantago ovata* husk) nutritional intervention, *Chemom. Intell. Lab. Syst.* 121 (2013) 1–8. doi:10.1016/j.chemolab.2012.11.015.
- [28] S. De Luca, M. De Filippis, R. Bucci, A.D. Magrì, A.L. Magrì, F. Marini, Characterization of the effects of different roasting conditions on coffee samples

- of different geographical origins by HPLC-DAD, NIR and chemometrics, *Microchem. J.* 129 (2016) 348–361. doi:10.1016/j.microc.2016.07.021.
- [29] F. Marini, D. de Beer, E. Joubert, B. Walczak, Analysis of variance of designed chromatographic data sets: The analysis of variance-target projection approach, *J. Chromatogr. A.* 1405 (2015) 94–102. doi:10.1016/j.chroma.2015.05.060.
- [30] R. Climaco-Pinto, A.S. Barros, N. Locquet, L. Schmidtke, D.N. Rutledge, Improving the detection of significant factors using ANOVA-PCA by selective reduction of residual variability, *Anal. Chim. Acta.* 653 (2009) 131–142. doi:10.1016/j.aca.2009.09.016.
- [31] F.T. Matsunaga, M. Rakocevic, J.D. Brancher, Modeling the 3D structure and rhythmic growth responses to environment in dioecious yerba-mate, *Ecol. Modell.* 290 (2014) 34–44. doi:10.1016/j.ecolmodel.2013.10.035.
- [32] C.I. Heck, E.G. De Mejia, Yerba mate tea (*Ilex paraguariensis*): A comprehensive review on chemistry, health implications, and technological considerations, *J. Food Sci.* 72 (2007) 138–151. doi:10.1111/j.1750-3841.2007.00535.x.
- [33] M.P. Arrieta, L. Peponi, D. López, M. Fernández-García, Recovery of yerba mate (*Ilex paraguariensis*) residue for the development of PLA-based bionanocomposite films, *Ind. Crops Prod.* 111 (2018) 317–328. doi:10.1016/j.indcrop.2017.10.042.
- [34] T.F.F. Da Silveira, A.D. Meinhart, T.C.L. De Souza, J. Teixeira Filho, H.T. Godoy, Phenolic compounds from yerba mate based beverages - A multivariate

- optimisation, *Food Chem.* 190 (2016) 1159–1167.
doi:10.1016/j.foodchem.2015.06.031.
- [35] J.D.P. Lima, A. Farah, B. King, T. De Paulis, P.R. Martin, Distribution of major chlorogenic acids and related compounds in brazilian green and toasted *Ilex paraguariensis* (Maté) leaves, *J. Agric. Food Chem.* 64 (2016) 2361–2370.
doi:10.1021/acs.jafc.6b00276.
- [36] B. Gullón, G. Eibes, M.T. Moreira, R. Herrera, J. Labidi, P. Gullón, Yerba mate waste: A sustainable resource of antioxidant compounds, *Ind. Crops Prod.* 113 (2018) 398–405. doi:10.1016/j.indcrop.2018.01.064.
- [37] A.B. Nicotra, R.L. Chazdon, R.A. Montgomery, Sexes show contrasting patterns of leaf and crown carbon gain in a dioecious rainforest shrub, *Am. J. Bot.* 90 (2003) 347–355. doi:10.3732/ajb.90.3.347.
- [38] M. Rakocevic, M.J.S. Medrado, S.F. Martim, E.D. Assad, Sexual dimorphism and seasonal changes of leaf gas exchange in the dioecious tree *Ilex paraguariensis* grown in two contrasted cultivation types, *Ann. Appl. Biol.* 154 (2009) 291–301.
doi:10.1111/j.1744-7348.2008.00298.x.
- [39] C.N.T. Frizon, G.A. Oliveira, C.A. Perussello, P.G. Peralta-Zamora, A.M.O. Camlofski, Ü.B. Rossa, R. Hoffmann-Ribani, Determination of total phenolic compounds in yerba mate (*Ilex paraguariensis*) combining near infrared spectroscopy (NIR) and multivariate analysis, *LWT - Food Sci. Technol.* 60 (2015) 795–801. doi:10.1016/j.lwt.2014.10.030.

- [40] C.I. Heck, M. Schmalko, E.G. De Mejia, Effect of growing and drying conditions on the phenolic composition of Mate teas (*Ilex paraguariensis*), *J. Agric. Food Chem.* 56 (2008) 8394–8403. doi:10.1021/jf801748s.
- [41] V.J. Barwick, Strategies for solvent selection - a literature review, *TrAC - Trends Anal. Chem.* 16 (1997) 293–309. doi:10.1016/S0165-9936(97)00039-3.
- [42] X. Li, Y. Zhang, Y. He, Rapid detection of talcum powder in tea using FT-IR spectroscopy coupled with chemometrics, *Sci. Rep.* 6 (2016) 1–8. doi:10.1038/srep30313.
- [43] M.J. Mohammed, F.A. Al-Bayati, Isolation, identification and purification of caffeine from *Coffea arabica* L. and *Camellia sinensis* L.: A combination antibacterial study, *Int. J. Green Pharm.* 3 (2009) 52–57. doi:10.4103/0973-8258.49375.
- [44] J. Ohnsmann, G. Quintás, S. Garrigues, M. De la Guardia, Determination of caffeine in tea samples by Fourier transform infrared spectrometry, *Anal. Bioanal. Chem.* 374 (2002) 561–565. doi:10.1007/s00216-002-1503-8.
- [45] I. Ścibisz, M. Reich, S. Bureau, B. Gouble, M. Causse, D. Bertrand, C.M.G.C. Renard, Mid-infrared spectroscopy as a tool for rapid determination of internal quality parameters in tomato, *Food Chem.* 125 (2011) 1390–1397. doi:10.1016/j.foodchem.2010.10.012.
- [46] A.S. Holt, E.E. Jacobs, Infra-Red absorption spectra of chlorophylls and derivatives, *Plant Physiol.* 30 (1955) 553–559. doi:10.1104/pp.30.6.553.

- [47] X. Li, R. Zhou, K. Xu, J. Xu, J. Jin, H. Fang, Y. He, Rapid determination of chlorophyll and pheophytin in green tea using fourier transform infrared spectroscopy, *Molecules*. 23 (2018) 1–13. doi:10.3390/molecules23051010.
- [48] P. Sanjay, K. Deepa, J. Madhavan, S. Senthil, Optical, spectral and photovoltaic characterization of natural dyes extracted from leaves of *Peltophorum pterocarpum* and *Acalypha amentacea* used as sensitizers for ZnO based dye sensitized solar cells, *Opt. Mater. (Amst)*. 83 (2018) 192–199. doi:10.1016/j.optmat.2018.06.011.
- [49] C. Eijkelhoff, J.P. Dekker, A routine method to determine the chlorophyll *a*, pheophytin *a* and β -carotene contents of isolated Photosystem II reaction center complexes, *Photosynth. Res.* 52 (1997) 69–73. doi:10.1023/A:1005834006985.
- [50] G.C. Coelho, M.F.G. Rachwal, R.A. Dedecek, G.R. Curcio, K. Nietsche, E.P. Schenkel, Effect of light intensity on methylxanthine contents of *Ilex paraguariensis* A. St. Hil., *Biochem. Syst. Ecol.* 35 (2007) 75–80. doi:10.1016/j.bse.2006.09.001.
- [51] Ü.B. Rossa, A.C. Angelo, J.Z. Mazuchowski, D.J. Westphalen, C.N.T. Frizon, C.E.N. Martins, Influência da luminosidade e fertilizantes nos teores de metilxantinas e compostos fenólicos em folhas de erva-mate, *Cienc. Florest.* 27 (2017) 1365–1374.
- [52] D.M. Da Croce, CARACTERÍSTICAS FÍSICO-QUÍMICAS DE EXTRATOS DE ERVA-MATE (*Ilex paraguariensis* St. Hil) NO ESTADO DE SANTA CATARINA, *Cienc. Florest.* 12 (2002) 107–113.

- [53] G.P. Scipioni, B. del V. Argüello, M.E. Schmalko, The Effect of Mg^{2+} , Cu^{2+} and Zn^{2+} Pre-treatment on the color of yerba maté (*Ilex paraguariensis*) leaves, *Brazilian Arch. Biol. Technol.* 53 (2010) 1497–1502.
- [54] M. Zarek, Seasonal fluctuations of photosynthetic pigments content in *Taxus baccata* needles, *Dendrobiology.* 76 (2016) 13–24.
- [55] L. Chen, S. Zhang, H. Zhao, H. Korpelainen, C. Li, Sex-related adaptive responses to interaction of drought and salinity in *Populus yunnanensis*, *Plant, Cell Environ.* 33 (2010) 1767–1778. doi:10.1111/j.1365-3040.2010.02182.x.
- [56] T.R. Randriamanana, L. Nybakken, A. Lavola, P.J. Aphalo, K. Nissinen, R. Julkunen-Tiitto, Sex-related differences in growth and carbon allocation to defence in *Populus tremula* as explained by current plant defence theories, *Tree Physiol.* 34 (2014) 471–487. doi:10.1093/treephys/tpu034.
- [57] G.G. Marcheafave, C.D. Tormena, E.D. Pauli, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Experimental mixture design solvent effects on pigment extraction and antioxidant activity from *Coffea arabica* L. leaves, *Microchem. J.* 146 (2019) 713–721. doi:10.1016/j.microc.2019.01.073.
- [58] P.H. Hatumura, G.S. de Oliveira, G.G. Marcheafave, M. Rakocevic, R.E. Bruns, I.S. Scarminio, A.E. Terrile, Chemometric Analysis of 1H NMR Fingerprints of *Coffea arabica* Green Bean Extracts Cultivated under Different Planting Densities, *Food Anal. Methods.* 11 (2018) 1906–1914. doi:10.1007/s12161-017-1104-y.

- [59] T.F.F. da Silveira, M. de A. Lima, A.D. Meinhart, G.G.C. Kuhnle, H.T. Godoy, Effect of solvent composition on the extraction of phenolic compounds and antioxidant capacity of bacaba juice (*Oenocarpus bacaba* Mart.), *Food Anal. Methods*. 13 (2020) 1119–1128. doi:10.1007/s12161-020-01726-2.
- [60] M.K. de A. Mendes, C.B. dos S. Oliveira, M.D.A. Veras, B.Q. Araújo, C. Dantas, M.H. Chaves, C.A. Lopes Júnior, E.C. Vieira, Application of multivariate optimization for the selective extraction of phenolic compounds in cashew nuts (*Anacardium occidentale* L.), *Talanta*. 205 (2019) 120100. doi:10.1016/j.talanta.2019.06.100.

CAPÍTULO II

Autenticação de cultivares de feijão carioca
(*Phaseolus vulgaris* L.) usando processamento
digital de imagens e ferramentas quimiométricas



Publicado em: *Food Chemistry*, Volume 364, Junho 2021, 130349.

RESUMO

A autenticação de feijão pode resultar em produtos de maior qualidade para o comércio. A análise discriminante por mínimos quadrados parciais (do inglês, *partial least squares discriminant analysis* - PLS-DA) foi aplicada a imagens digitais a fim de desenvolver uma metodologia que permite a discriminação não destrutiva de três cultivares de *Phaseolus vulgaris* L. (Agro ANfc9, IPR-Andorinha e IPR-Sabiá) com diferentes características tecnológicas. A análise de componentes principais (PCA) resultou na separação dessas cultivares, mas com certa sobreposição. A análise supervisionada mostrou que três modelos PLS1-DA, um modelo para duas cultivares, foi moderadamente melhor do que o tratamento simultâneo de todas as três cultivares (PLS2-DA). O teste de permutação avaliou a significância estatística dos modelos PLS-DA. Melhores classificações foram encontradas para as cultivares Agro ANfc9 e IPR-Sabiá do que para IPR-Andorinha. O modelo Agro ANfc9 – IPR-Sabiá classificou corretamente 100% das duas classes de feijão nos conjuntos de treinamento e teste. Essa estratégia analítica é rápida, econômica, ambientalmente sustentável e pode ser aplicada no controle de qualidade do feijão, ajudando na autenticidade do cultivo para o comércio garantindo um produto com características como menor tempo de cozimento, maior resistência a doenças, composição química mais desejável e maior vida útil.

1 INTRODUÇÃO

O feijão comum (*Phaseolus vulgaris* L.) é uma cultura mundialmente significativa e uma leguminosa importante em termos de rendimento total e área cultivada, sendo os principais produtores Mianmar (5,8 Mt), Índia (5,3 Mt), Brasil (3,0 Mt), Estados Unidos (932,2 t) e México (879,4 t) (Gioia et al., 2019; OECD, 2019; Silva et al., 2020). Seu maior consumo ocorre em países subdesenvolvidos principalmente por causa de seu alto valor nutricional e característica de saciedade devido a concentrações significativas de proteínas, carboidratos, fibras, vitaminas e minerais (Mavromatis et al., 2012; OECD, 2019; Silochi et al., 2016).

Apesar de ser consumido principalmente na forma de feijão seco e enlatado, o feijão pode ser usado como diferentes tipos de ingredientes para a produção de alimentos com melhor valor nutricional, como farinha, pão, bolos, tortilhas, entre outros (Mavromatis et al., 2012; OECD, 2019). Porém, o consumidor exige qualidade do produto e busca maior diversificação nas gôndolas dos supermercados, bem como produtos que sejam rápidos e fáceis de preparar, respeitando as características sensoriais esperadas (Pedrão & Coró, 1999).

A qualidade do feijão é determinada por dois fatores: valor nutricional e características tecnológicas. Esse último depende da variedade de cultivares e determina a aceitação pelo consumidor de atributos físicos como cor, brilho e textura que estão diretamente relacionados ao tempo de cozimento (Silochi et al., 2016). No entanto, muitas das cultivares são de difícil identificação devido ao grande número de características semelhantes obtidas no cruzamento de genótipos dos feijões de interesse (da Costa et al., 1979). Quando as naturezas morfológicas são visualmente distinguíveis, a identificação das cultivares torna-se mais fácil (Ferreira et al., 2011). Contudo, quando existem

pequenas variações na cor, forma do hilo, tonalidade e brilho do tegumento, os métodos de inspeção tornam-se insuficientes para a identificação segura de cultivares. Isso é importante para os consumidores e a indústria, a fim de obter informações precisas, completas e verdadeiras sobre os alimentos vendidos e consumidos (Böhme et al., 2019). A autenticação de alimentos inclui vários tópicos, como identificação de tipo e/ou variedade, que muitas vezes exige o uso de estratégias analíticas sofisticadas e com alto custo devido ao fluxo dinâmico de produtos e trocas comerciais. Isso exige o desenvolvimento de ferramentas de autenticação que sejam rápidas, seguras, de baixo custo e ecológicas, além de serem precisas (Böhme et al., 2019).

O uso de imagens digitais fornece uma forma de substituir a inspeção visual humana, que é suscetível a erros devido ao seu caráter subjetivo e às influências sofridas por fatores ambientais externos (Byrne et al., 2000). Os sistemas de cores usados com mais frequência definem um espaço de coordenadas tridimensional, para o qual cada cor é representada por um único ponto com valores de vermelho-verde-azul (RGB), matiz-saturação-intensidade (HSI) e tons de cinza (Costa et al., 2015). Ao longo dos anos, a aplicação de imagens digitais para identificar adulteração, reconhecimento de padrões e qualidade alimentar tem sido relatada com sucesso (Antonelli et al., 2004; Costa et al., 2015; Ferreira et al., 2018; Fulgêncio et al., 2020; Grasel et al., 2016, 2017; Santos et al., 2012). Seu uso, quando associado a ferramentas quimiométricas, permite análises econômicas, rápidas e não destrutivas, sem exigir o uso de equipamentos sofisticados, pois as imagens podem ser obtidas em câmeras digitais, smartphones, scanners e webcams.

As estratégias de processamento para análise de imagem digital podem variar amplamente, de regressões univariadas simples a modelos multivariados. A abordagem de análise discriminante pode ser realizada por análise de regressão com uma

matriz de indicador dando as classes de observações no conjunto de treinamento (Chevallier et al., 2006). Uma das principais e mais populares estratégias de análise de imagens digitais é o uso de calibração multivariada, principalmente pela aplicação de mínimos quadrados parciais (PLS) (Fulgêncio et al., 2020). O algoritmo PLS é a chave para o desenvolvimento da conhecida análise discriminante por mínimos quadrados parciais (PLS-DA), em que classes rotuladas são usadas como um vetor y dependente (Milanez & Pontes, 2015; Tormena et al., 2019). Aqui o PLS-DA é aplicado ao conjunto de dados para classificação da amostra, além de reduzir a complexidade da modelagem matemática (Marcheafave et al., 2020).

O feijão comum possui grande diversidade de cores, sendo que o tipo carioca tem o bege como cor principal e o marrom como cor secundária, apresentando-se em forma de estrias. A clareza dos grãos está diretamente relacionada à preferência do consumidor do produto e ao valor comercial, sendo também um parâmetro promissor de grande interesse para autenticação (Antonelli et al., 2004; Possobom et al., 2015). Até onde sabemos, a técnica de imagem digital ainda não foi explorada para a autenticação de grãos de diferentes cultivares. Aqui é investigado seu uso como uma forma rápida e precisa de classificar o feijão comum para garantir sua qualidade com características como menor tempo de cozimento, maior resistência a doenças, composição química mais desejável e maior vida útil. Dessa forma, o uso de imagens digitais acopladas ao PLS-DA possibilita a autenticação de três cultivares de *Phaseolus vulgaris* L. com diferentes características tecnológicas evitando fraudes tanto na produção quanto na comercialização.

2 MATERIAIS E MÉTODOS

2.1. AMOSTRAS

Três cultivares de feijão comum (*Phaseolus vulgaris* L.) do tipo carioca, IPR-Andorinha e IPR-Sabiá, ambas do IDR-Paraná, e Agro ANfc9 da Agro Norte, desenvolvidas por dois centros de pesquisa (Instituto de Desenvolvimento Rural do Paraná - Londrina - Paraná - Brasil (IDR-Paraná) e Agro Norte Pesquisa e Sementes - Sinop - Mato Grosso - Brasil) foram utilizados neste trabalho. Eles foram escolhidos por suas diferentes características tecnológicas e agronômicas, como tempo de cozimento, vida útil, rendimento agrônômico, resistência a doenças e duração do ciclo de produção.

As amostras foram cultivadas e fornecidas pelo IDR-Paraná em 2019 ($-23^{\circ}18'37''\text{S}$, $51^{\circ}09'46''\text{O}$, 585 m de altitude). Feijões com quaisquer defeitos foram removidos para padronizar as amostras. Os grãos sem defeitos foram armazenados em sacos de papel em ultrafreezer a $-60\text{ }^{\circ}\text{C}$. Para a análise, foram utilizados 60 grãos diferentes de cada cultivar.

2.2. SISTEMA DE AQUISIÇÃO DE IMAGENS DIGITAIS

O sistema de aquisição é mostrado esquematicamente na Fig. 1. Esse sistema é composto por uma caixa de fibra de média densidade (MDF) de $20 \times 40 \times 20$ cm com uma abertura de porta, usada para isolamento completo da luz. A cor das amostras está ligada à distribuição espectral da luz; portanto, a captura da imagem depende da fonte de iluminação (Diniz, 2020). Dessa forma, o interior da caixa foi equipado com duas fitas de LED de três metros com alimentação de 12 V e um total de 600 LEDs, uma *webcam full HD 1080p* (Logitech Vidro Full HD C920 de 15 MP, com conexão USB) posicionada a 180° e 4 cm em relação ao compartimento da amostra. A caixa estava sempre fechada

durante as aquisições das imagens.

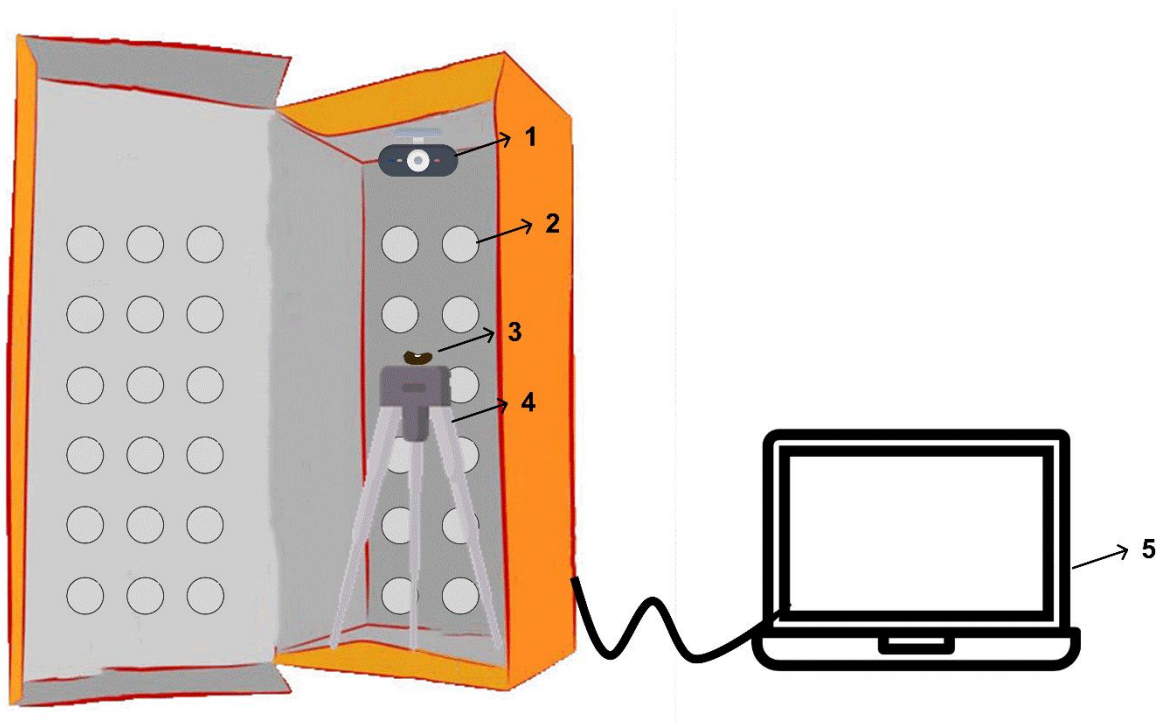


Figura 1. Esquema de aquisição de imagens digitais: (1) *Webcam*; (2) Fonte de luz LED; (3) Compartimento da amostra; (4) Suporte; (5) Computador para aquisição dos dados.

O sistema mostrado na Figura 1 foi conectado a um computador portátil utilizando o visualizador Picture Manager para a obtenção das imagens que foram salvas no formato JPG, com dimensões de 1280×720 pixels e 96 dpi. Uma região do feijão foi selecionada e cortada, definindo-a como a região de interesse (ROI) com 50×50 pixels (Figura 2). Isso foi feito para todas as imagens. Essa área foi usada para extrair os histogramas de frequência dos níveis de cor RGB padrão, tons de cinza e HSI. Cada componente de cor é representado por 8 bits, assumindo valores de 0 a 255 (totalizando 256 variáveis). Portanto, o vetor do histograma (\mathbf{x}) de cada amostra, considerando os sete componentes da cor adotados neste estudo, é de 1×1792 . Foram investigadas três classes de feijão do tipo carioca, Agro ANfc9, IPR-Andorinha e IPR-Sabiá. Para cada classe, 60 imagens foram adquiridas e organizadas em uma matriz 60×1792 .

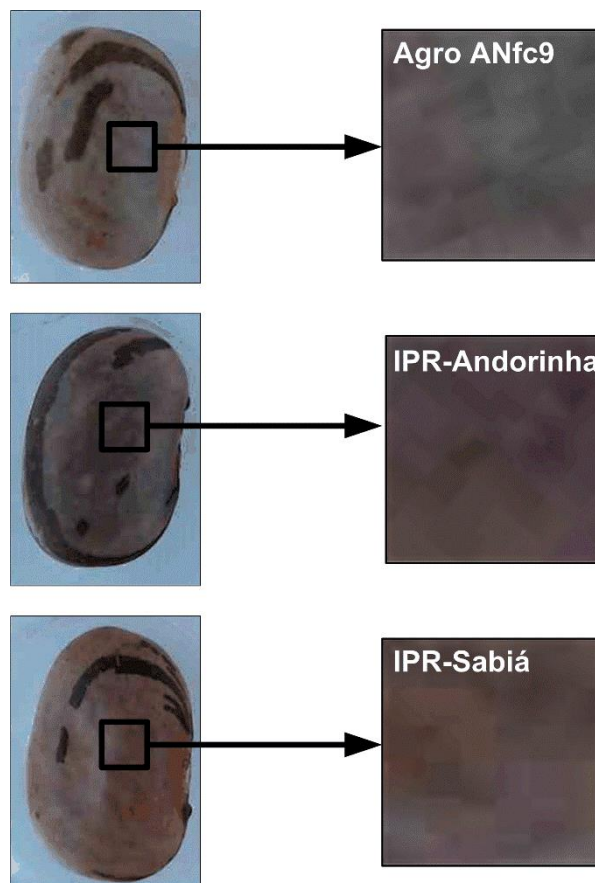


Figura 2. Região de interesse (ROI) utilizada a partir de imagens digitais adquiridas para as três classes de cultivares de feijão: Agro ANfc9, IPR-Andorinha e IPR-Sabiá.

2.3. ANÁLISE MULTIVARIADA

A análise de componentes principais (PCA) é uma ferramenta exploratória de aprendizagem multivariada não supervisionada que transforma dados laboratoriais multivariados em dimensões menores, chamadas de componentes principais (PCs). As PCs ajudam a identificar a formação de padrões de amostra (score) devido às diferenças e semelhanças variáveis. *Outliers* também podem ser encontrados (Marini, 2013). A matriz 60×1792 foi centrada na média e então explorada por PCA.

O PLS-DA é baseado em um algoritmo de regressão PLS e tenta construir modelos que possam maximizar a separação entre classes de objetos (Suhandy

& Yulia, 2017). Seu objetivo é construir modelos de classificação que possam ser aplicados para previsões futuras (Santana et al., 2016). No PLS-DA a matriz \mathbf{X} (que contém as respostas das propriedades de interesse, neste caso, os componentes de cor) está relacionada a outra matriz, \mathbf{Y} , ou vetor \mathbf{y} , contendo informações sobre as classes da amostra, como tal sendo um método supervisionado de reconhecimento de padrões (Pereira et al., 2018; Valderrama & Valderrama, 2016). Quando os dados contêm duas classes, a variável \mathbf{y} será composta de um único vetor que é codificado com 0 para todas as amostras da primeira classe e 1 para todas as amostras da segunda classe. Se o conjunto de dados contiver três classes, elas serão codificadas em binários em 3 variáveis, *i.e.*, com a matriz \mathbf{Y} como [1,0,0] para amostras da classe 1; [0,1,0] para amostras da classe 2; e [0,0,1] para amostras da classe 3. Portanto, no caso de apenas duas classes de cada vez serem modeladas, o método é denominado PLS1-DA comum, enquanto o tratamento simultâneo de várias classes é modelado com o método PLS2-DA (Bassbasi et al., 2014).

O modelo PLS-DA é baseado em um limiar de discriminação Bayesiano, que primeiramente estima as probabilidades para depois discriminar as amostras. Portanto, é necessário encontrar um valor limite para as amostras do conjunto de treinamento e isso é melhor feito dividindo essas classes a fim de obter a menor probabilidade de classificações falsas de previsões futuras (Pereira et al., 2018; Tormena et al., 2019). Então, valores acima desse limite indicam que as amostras pertencem à classe modelada, enquanto valores mais baixos indicam que as amostras não pertencem a essa classe (Pereira et al., 2018; Valderrama & Valderrama, 2016).

Após realizar PCA o modelo PLS-DA foi construído. Primeiro, os componentes da cor foram sistematicamente divididos em dois subconjuntos usando o algoritmo Duplex (Snee, 1977), separadamente para cada classe. Um conjunto de treinamento, representando 70% das amostras, foi utilizado para modelagem e validação

interna, ou seja, 42 amostras para cada classe. O conjunto de teste contendo 30% ou 18 amostras para cada classe, foi utilizado para validação externa do modelo desenvolvido. Os dados foram centrados na média e o número de variáveis latentes (LV) foi escolhido com base no menor erro de classificação de validação cruzada (CVCE) (Ballabio & Consonni, 2013; Santos et al., 2019) obtido pela veneziana da validação cruzada. O modelo foi otimizado para detecção de *outlier*, baseado nos resíduos T^2 e Q de Hotelling, ambos com nível de confiança de 95% (Ballabio & Consonni, 2013). A avaliação da classificação foi realizada por parâmetros de desempenho de sensibilidade, especificidade e precisão (Ballabio & Consonni, 2013; Tormena et al., 2019). Além disso, foi aplicado o teste de permutação (Szymanska et al., 2012), para validação da discriminação entre classes. Nesse caso, os rótulos das amostras são permutados aleatoriamente e um novo modelo de classificação é calculado, assumindo que não há diferença entre os grupos formados aleatoriamente (Szymanska et al., 2012). Para isso, 200 iterações foram realizadas comparando modelos permutados com não permutados, levando em consideração a auto-predição e a validação cruzada.

2.4. ANÁLISE ESTATÍSTICA E *SOFTWARE*

O *software* Matlab 2016b® (MathWorks, Natick, MA, EUA) foi utilizado para converter as imagens originais em histogramas das cores e para a aplicação o PLS toolbox 8.7® (Eigenvector Research) no tratamento e análise quimiométrica.

3 RESULTADOS E DISCUSSÕES

Os programas de melhoramento do feijão visam melhorar as características específicas de cultivares sem comprometer outras (Kelly, 2010). Eles devem encontrar várias características que sejam do interesse do produtor, empresário e consumidor. O produtor busca grãos com alto rendimento, alta resistência a doenças e baixo dano mecânico na colheita e manuseio, enquanto o empresário busca um produto de alta qualidade, sem danos externos, com bom aspecto visual e alto índice de hidratação. Ao mesmo tempo, o consumidor exige um produto que seja fácil de preparar, mas que mantenha as características sensoriais esperadas (Kelly, 2010; Pedrão & Coró, 1999). Combinar todas essas características diversas em uma única cultivar não é fácil. Portanto, a autenticação do feijão é importante para obter produtos com as qualidades necessárias para produtores, vendedores e consumidores. Nesta pesquisa, três cultivares foram escolhidas para sua autenticação levando em consideração suas respectivas características, como tempo de cocção, vida útil, resistência a doenças ou tempo de ciclo de produção.

3.1. ANÁLISE EXPLORATÓRIA

A investigação do ROI mostra que é difícil distinguir entre as três classes de feijão, no caso Agro ANfc9, IPR-Andorinha e IPR-Sabiá, devido à semelhança de suas cores. Como tal, a inspeção visual pode levar a resultados errôneos (Figura 2). No entanto, alguma variação na cor e intensidade deve existir nos grãos. Como as imagens são caracterizadas por uma textura estocástica natural da imagem, os histogramas de cores descrevem a distribuição estatística dos pixels em função do componente de cor registrado, e não de propriedades físico-químicas (Diniz et al., 2012). Por esse motivo, foi realizada uma análise exploratória para examinar os agrupamentos naturais das 180

amostras de feijão de acordo com seus cultivares, utilizando PCA e contemplando todos os componentes de cor adotados neste estudo - RGB, tons de cinza e HSI. A Figura 3 exibe os histogramas para o espaço de cores na escala de cinza, RGB e HSI, para todas as amostras das três cultivares de *Phaseolus vulgaris* L.

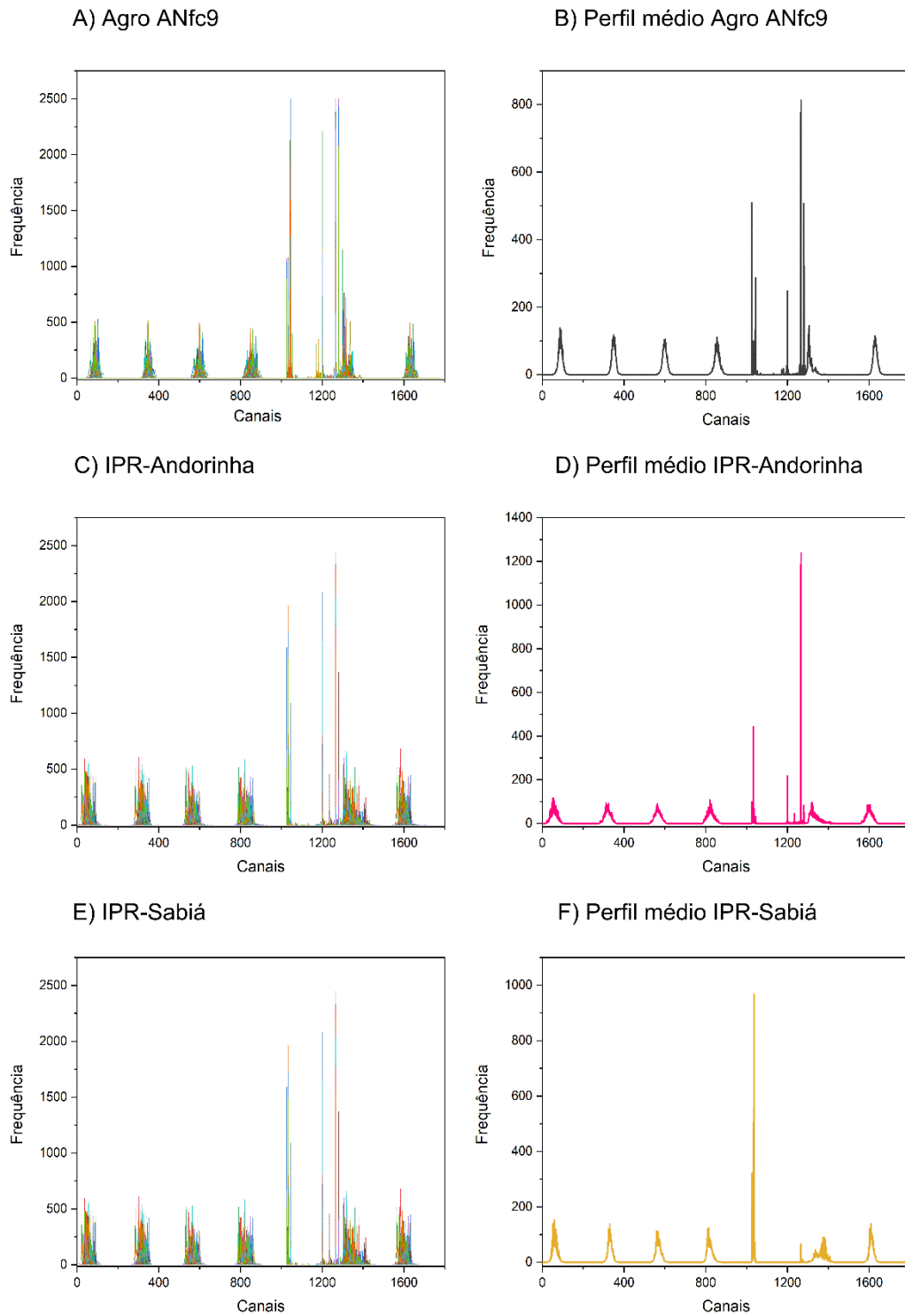


Figura 3. Histogramas para o espaço de cores na escala de cinza, RGB e HSI para todas as amostras dos três cultivares de feijão e suas médias. A) Agro ANfc9; B) Perfil médio do Agro ANfc9; C) IPR-Andorinha; D) Perfil médio do IPR-Andorinha; E) IPR-Sabiá; e F) Perfil médio do IPR-Sabiá.

A Figura 4 mostra um gráfico de dispersão bidimensional de escores para as primeiras duas PCs. O gráfico PC1×PC2 descreve a maior variação dos dados do que qualquer outro par de componentes, sendo responsável por 26,19% na PC1 e 16,10% na PC2 da variância total. Como pode ser visto, há uma separação entre os grupos correspondentes às três classes de grãos, com algum grau de sobreposição. Isso pode ser esperado, pois apesar de pertencerem a diferentes cultivares e possuírem características tecnológicas distintas, existe uma grande similaridade entre os grãos do tipo carioca. Portanto, a PLS-DA, uma técnica supervisionada, foi utilizada para classificar as amostras de acordo com suas cultivares na aplicação do controle de qualidade do feijão.

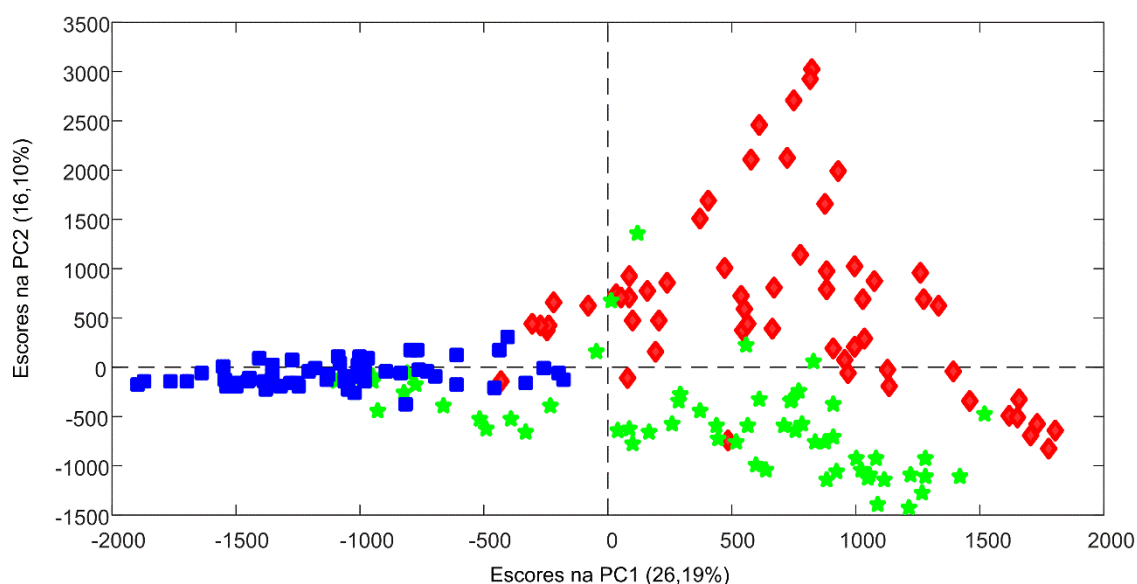


Figura 4. Gráfico dos escores da PCA obtidos a partir dos histogramas RGB, tons de cinza e HSI para todas as amostras de feijão: Agro ANfc9 (♦), IPR-Andorinha (★) e IPR-Sabiá (■).

3.2. CLASSIFICAÇÃO

Para construir o modelo PLS2-DA para tratar as três classes simultaneamente, todo o conjunto de dados foi dividido em 126 amostras para o conjunto de treinamento (42 para cada classe) e 54 amostras para o conjunto de teste (18 para cada classe), conforme mencionado na Seção 2.4. O número ideal de variáveis latentes (LV) foi escolhido com base nos menores valores de CVCE. A identificação de *outlier* no modelo foi realizada com Hotelling T^2 versus resíduos Q. Uma vez que todas as amostras apresentaram, simultaneamente, resíduos de Hotelling T^2 e Q abaixo dos limites de tolerância, não foram identificados *outliers*, como pode ser visto na Figura 5A. O melhor modelo PLS2-DA continha 4 LV, respondendo por 59,29% da variância no bloco **X** e 77,41% no bloco **Y**. O modelo otimizado foi então validado com o conjunto de teste externo. Os resultados são mostrados na Figura 6. A linha horizontal vermelha é o limiar bayesiano que define os limites da classe.

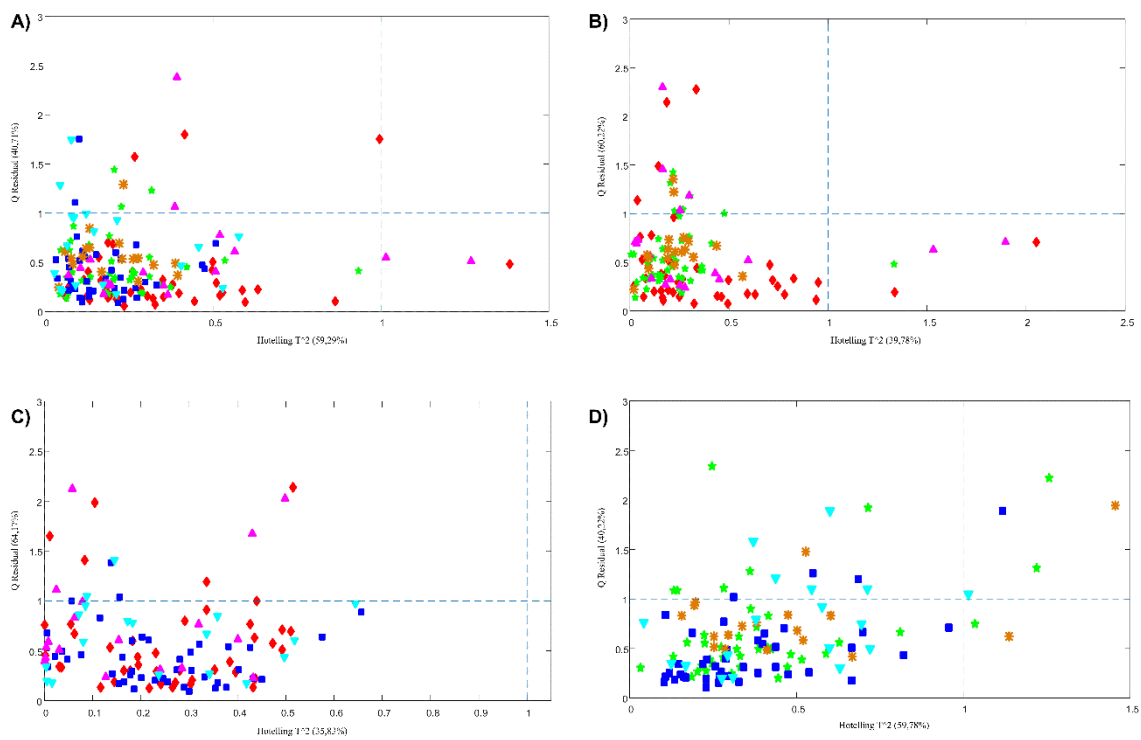


Figura 5. Resíduos Q *versus* Hotelling T^2 para o modelo PLS-DA de classificação de classes de cultivares de feijão. A) Todas as classes; B) Agro ANfc9 versus IPR-Andorinha; C) Agro ANfc9 versus IPR-Sabiá; D) IPR-Sabiá versus IPR-Andorinha. Predição Agro ANfc9 (◆), validação Agro ANfc9 (▲), predição IPR-Andorinha (★), validação IPR-Andorinha (*), predição IPR-Sabiá (■) e validação IPR-Sabiá (▼).

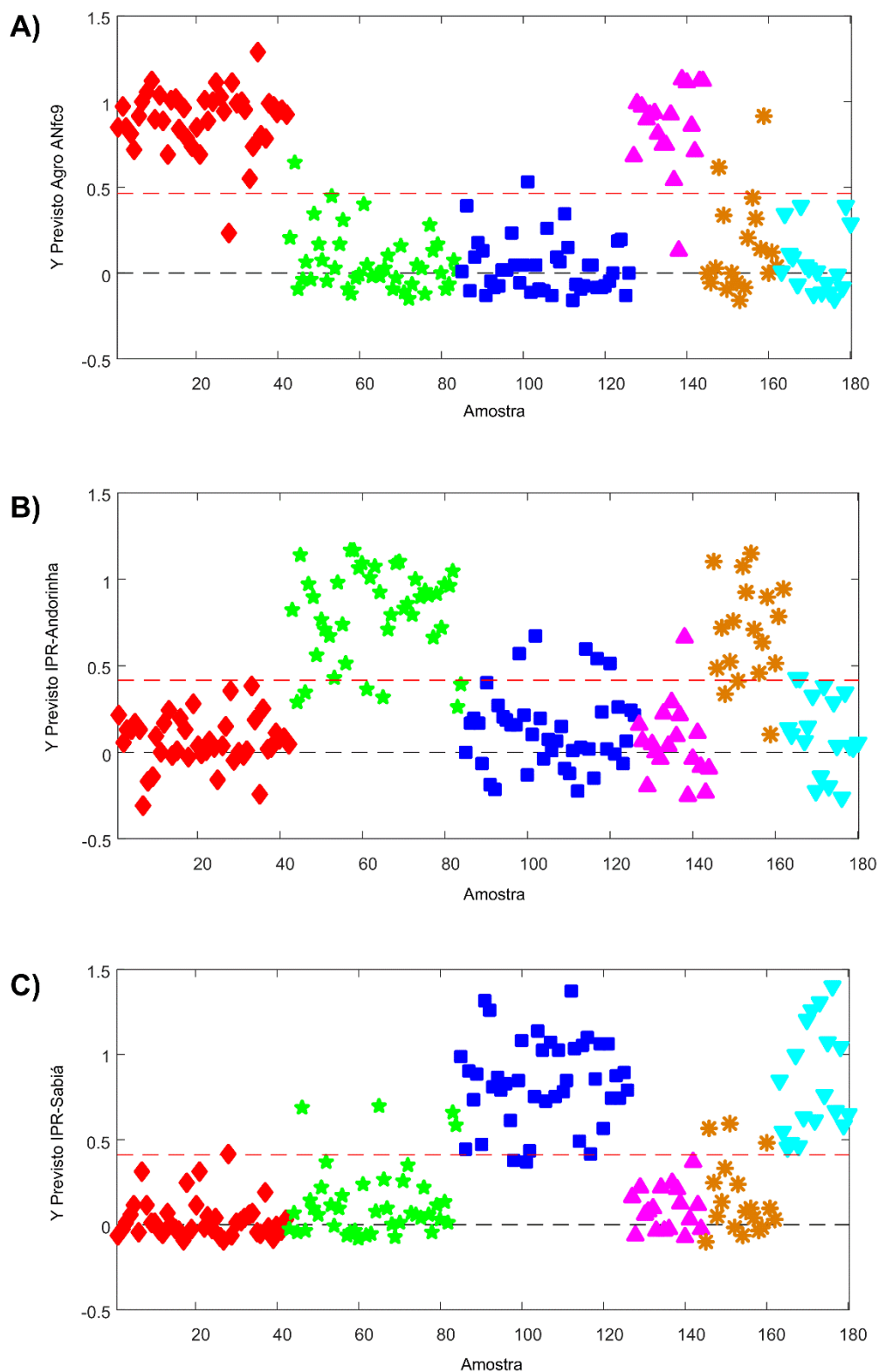


Figura 6. Valores de Y preditos para o modelo PLS2-DA para a classificação das classes dos cultivares Agro ANfc9, IPR-Andorinha e IPR-Sabiá. Predição Agro ANfc9 (◆), validação Agro ANfc9 (▲), predição IPR-Andorinha (★), validação IPR-Andorinha (*),

predição IPR-Sabiá (■) e validação IPR-Sabiá (▼).

A Tabela 1 mostra os parâmetros de qualidade de precisão, sensibilidade e especificidade para modelos PLS2-DA. A sensibilidade de validação do modelo, dada pelo número de amostras corretamente classificadas para pertencer a uma classe dividido pelo número de amostras realmente daquela classe (Valderrama & Valderrama, 2016), foi de 100% para IPR-Sabiá. Para a classe Agro ANfc9, a sensibilidade foi igual a 94,4%, ou seja, uma amostra não estava prevista em sua respectiva classe, tornando-a uma amostra falso negativa. Os feijões IPR-Andorinha apresentaram sensibilidade de predição de 83,3%; isso porque três amostras dessa classe não foram previstas como IPR-Andorinha. A especificidade refere-se à predição incorreta de amostras teste de outras classes na classe avaliada (Valderrama & Valderrama, 2016). A classe IPR-Sabiá apresentou especificidade de 91,7%, pois três amostras de IPR-Andorinha foram incorretamente preditas como IPR-Sabiá. Para o IPR-Andorinha, a especificidade foi de 91,7%, devido a uma amostra de Agro ANfc9 e duas amostras de IPR-Sabiá preditas como IPR-Andorinha. Finalmente, para a classe Agro ANfc9 a especificidade foi de 94,4%, com duas amostras IPR-Andorinha incorretamente previstas para estarem nesta classe. Os modelos de predição apresentaram boa precisão, ou seja, as proporções de classificações corretas, independente da classe, foram iguais a 94,4% para as classes Agro ANfc9 e IPR-Sabiá e 88,9% para a classe IPR-Andorinha. Os modelos PLS2-DA foram submetidos a testes de permutação. Os resultados indicam que esses modelos são significativamente diferentes ao nível de 95% dos não permutados. Também não são super ajustados porque os valores de Wilcoxon (teste de classificação com sinais de Wilcoxon), teste de sinais (teste de classificação com sinais em pares) e teste *t* de Rand (teste *t* de randomização) são menores que 0,05 (Tabela 2).

Tabela 1. Parâmetros de classificação e predição de qualidade para os feijões dos cultivares Agro ANfc9, IPR-Andorinha e IPR-Sabiá pelos modelos PLS2-DA e PLS1-DA comum.

Modelo	PLS2-DA			PLS1-DA					
	<i>Agro ANfc9</i>	<i>IPR-Andorinha</i>	<i>IPR-Sabiá</i>	1		2		3	
				<i>Agro ANfc9</i>	<i>IPR-Andorinha</i>	<i>Agro ANfc9</i>	<i>IPR-Sabiá</i>	<i>IPR-Andorinha</i>	<i>IPR-Sabiá</i>
Sensibilidade (Treinamento)	97,6	85,7	95,2	97,6	97,6	97,6	100,0	90,0	97,6
Especificidade (Treinamento)	97,6	94,0	94,0	97,6	97,6	100,0	97,6	97,6	90,0
Precisão (Treinamento)	97,6	91,3	94,4	97,6		98,8		93,8	
Sensibilidade (Teste)	94,4	83,3	100,0	94,4	88,9	100,0	100,0	88,2	94,1
Especificidade (Teste)	94,4	91,7	91,7	88,9	94,4	100,0	100,0	94,1	88,2
Precisão (Teste)	94,4	88,9	94,4	91,7		100,0		91,2	
RMSEC	0,1798	0,2536	0,2323	0,1812		0,1875		0,2195	
RMSECV	0,1939	0,2771	0,2563	0,2058		0,1974		0,3344	
Variáveis Latentes		4		3		1		4	

Tabela 2. Parâmetros do teste de permutação obtidos para o modelo PLS2-DA.

	Wilcoxon	Teste de sinais	Teste <i>t</i> de Rand
Agro ANfc9			
Auto-predição (RMSEC*)	0,000	0,000	0,005
Validação cruzada (RMSECV**)	0,000	0,000	0,005
IPR-Andorinha			
Auto-predição (RMSEC)	0,000	0,000	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,005
IPR-Sabiá			
Auto-predição (RMSEC)	0,000	0,000	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,005

*Raiz quadrada do erro quadrático médio de calibração do inglês, *root mean square error of calibration* (RMSEC).

**Raiz quadrada do erro quadrático médio de validação cruzada do inglês, *root mean square error of cross-validation* (RMSECV).

Embora o modelo PLS2-DA apresente bons resultados entre as classes dessas cultivares de feijão, modelos comuns PLS1-DA foram construídos para avaliar se haveria ganho de qualidade na predição de grãos da cultivar IPR-Andorinha, na qual o modelo PLS2-DA forneceu os valores dos parâmetros de qualidade mais baixos. Assim, três modelos de autenticação binária para feijão do tipo carioca foram construídos para cada um dos três pares de cultivares, 1) Agro ANfc9 – IPR-Andorinha; 2) Agro ANfc9 – IPR-Sabiá; e 3) IPR-Andorinha – IPR-Sabiá (Figura 7). Cada conjunto binário foi submetido à detecção de *outlier* por meio da análise simultânea dos resíduos T^2 e Q de Hotelling. Para o modelo IPR-Andorinha – IPR-Sabiá, um total de cinco amostras (3 para IPR-Andorinha e 2 para IPR-Sabiá) foram detectadas como *outliers* (Figura 5D), correspondendo a 5,9% das amostras originais. Não foram detectados *outliers* para os modelos Agro ANfc9 – IPR-Andorinha e Agro ANfc9 – IPR-Sabiá (Figura 5B e C).

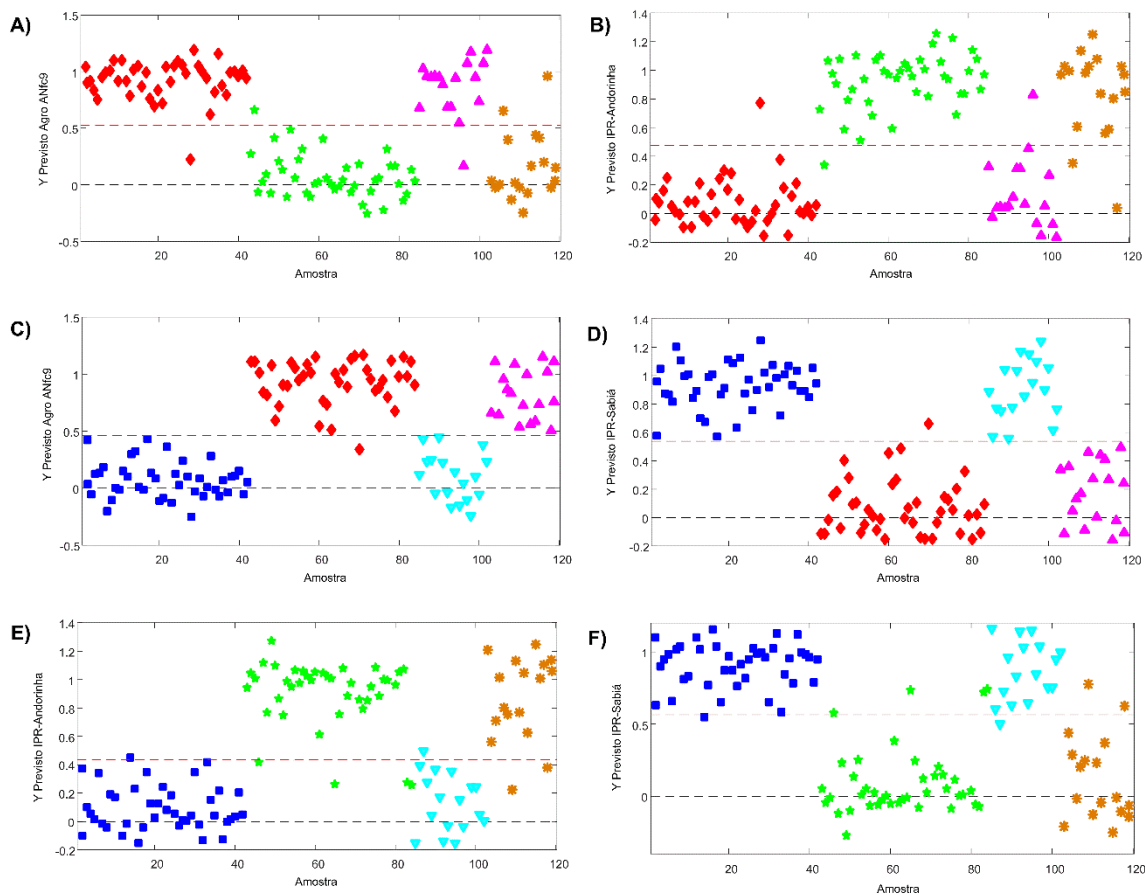


Figura 7. Valores de y e previstos para os modelos binários PLS1-DA: Agro ANfc9 – IPR-Andorinha (A) Agro ANfc9 classe e (B) IPR-Andorinha classe; Agro ANfc9 – IPR-Sabiá (C) classe Agro ANfc9 e (D) classe IPR-Sabiá; e IPR-Andorinha – IPR-Sabiá (E) classe IPR-Andorinha e (F) classe IPR-Sabiá. Predição Agro ANfc9 (♦), validação Agro ANfc9 (▲), predição IPR-Andorinha (★), validação IPR-Andorinha (*), predição IPR-Sabiá (■) e validação IPR-Sabiá (▼).

A Tabela 1 mostra que os resultados dos modelos de predição PLS1-DA são um pouco melhores que os do modelo PLS2-DA, em relação à classe IPR-Andorinha. Ou seja, para o modelo PLS1-DA construído com as classes Agro ANfc9 e IPR-Andorinha, o IPR-Andorinha tem uma sensibilidade de 88,9%, devido a duas de suas amostras não serem classificadas como tal, e a especificidade de 94,4% porque uma amostra do Agro ANfc9 foi predita incorretamente como IPR-Andorinha. Para o modelo

PLS1-DA construído com classes IPR-Andorinha – IPR-Sabiá, o IPR-Andorinha tem sensibilidade de 88,2% (duas amostras não foram previstas nesta classe) e especificidade de 94,1% (uma amostra IPR-Sabiá classificada como IPR-Andorinha). Portanto, esses resultados de predição para o modelo PLS1-DA foram ligeiramente maiores para as amostras da classe IPR-Andorinha em comparação aos valores de sensibilidade de 83,3% e especificidade de 91,7%, calculados pelo PLS2-DA. Além disso, o modelo do par Agro ANfc9 – IPR-Sabiá foi capaz de classificar corretamente as duas classes de feijão e não apresentou erro na predição das amostras do conjunto de teste, com sensibilidade e especificidade iguais a 100% (Tabela 1). Esses modelos PLS-DA construídos também foram submetidos a testes de permutação, e os resultados indicam que são significativamente diferentes dos não permutados ao nível de 95% (Tabela 3) e não apresentam sobreajuste.

Tabela 3. Parâmetros do teste de permutação obtidos para o modelo PLS1-DA.

	Wilcoxon	Teste de sinais	Teste <i>t</i> de Rand
Agro ANfc9			
Auto-predição (RMSEC*)	0,000	0,000	0,005
Validação cruzada (RMSECV**)	0,000	0,000	0,005
IPR-Andorinha			
Auto-predição (RMSEC)	0,000	0,000	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,005
Agro ANfc9			
Auto-predição (RMSEC)	0,000	0,000	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,005
IPR-Sabiá			

Auto-predição (RMSEC)	0,000	0,000	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,005
<hr/>			
IPR-Andorinha			
Auto-predição (RMSEC)	0,000	0,000	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,005
IPR-Sabiá			
Auto-predição (RMSEC)	0,000	0,000	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,005

*Raiz quadrada do erro quadrático médio de calibração do inglês, *root mean square error of calibration* (RMSEC).

**Raiz quadrada do erro quadrático médio de validação cruzada do inglês, *root mean square error of cross-validation* (RMSECV).

Em geral, os modelos PLS1-DA, ou seja, modelos binários para as três classes de feijão, são ligeiramente melhores do que o modelo PLS2-DA. Assim, o uso de imagens digitais associadas à ferramenta quimiométrica PLS-DA apresentou bons resultados para a classificação e predição de amostras de feijão. Apesar de serem visualmente muito semelhantes, os grãos avaliados neste estudo apresentam propriedades desejáveis diferentes. Portanto, essa metodologia pode ser aplicada no controle de qualidade do feijão, quanto à autenticidade de cultivares. As futuras aplicações de modelagem devem considerar fatores incontrolláveis como maturidade do feijão e diferentes períodos de colheita, entre outros.

A cultivar IPR-Sabiá apresenta grandes vantagens para o cultivo do feijão. As plantas estão em posição vertical, o que permite uma colheita mecanizada eficiente. Essa cultivar também apresenta boa resistência ao armazenamento e às principais doenças ativas do feijão. O ciclo médio de produção desse grão é de cerca de 87 dias. Além disso, é uma cultura que possui um tempo de cozimento de cerca de 20 minutos apenas, além de apresentar boa eficiência de colheita (Cittolin Sementes, 2020). A cultivar IPR-Andorinha apresenta potencial promissor de produtividade de colheita

tendo como principal vantagem o seu ciclo curto de 73 dias, o que possibilita o cultivo em diferentes sistemas de produção. Essa cultivar é resistente a algumas das principais doenças ativas do feijão e é moderadamente resistente a fatores abióticos. Sua qualidade tecnológica industrial e nutricional é indicada pelo curto tempo de cocção, cerca de 18 minutos, e alto percentual de proteína, em torno de 24% (IDR-Paraná, 2013). A cultivar Agro ANfc9 é a única não desenvolvida pelo IDR-Paraná. Entre suas principais características estão o alto desempenho agrônômico e o alto prazo de validade. Essa última característica afeta diretamente os consumidores desde o armazenamento, assim como a identidade intrínseca da semente, desencadeia reações químicas que alteram a cor e geram endurecimento (Siqueira et al., 2016). Por essa razão, a longa vida útil está ligada à não suscetibilidade do grão ao escurecimento e endurecimento. Por outro lado, a cultivar Agro ANfc9 tem ciclo de produção de 94 dias e apresenta resistência apenas a uma doença específica do feijoeiro, sendo, portanto, suscetível às demais (Agro Norte Pesquisas e Sementes, 2020).

Com a evolução da tecnologia e, conseqüentemente, com a alta tecnificação da agricultura, cada vez mais produtos autenticados para cultivares estão sendo demandados. Devem possuir, no mínimo, certificados que apresentem suas distinções qualitativas (estéticas, nutricionais, ambientais) garantindo ao consumidor a aquisição do produto de sua preferência (Pessoa et al., 2002). Os modelos aqui investigados foram capazes de autenticar as diferentes cultivares de feijão de forma rápida e não destrutiva garantindo ao consumidor a compra de produtos certificados com a qualidade exigida pelos produtores e consumidores.

4 CONCLUSÃO

Este estudo demonstra o grande potencial do uso de imagens digitais com seus histogramas de cores para autenticação de feijão *Phaseolus vulgaris* L. Essa análise qualitativa associada aos modelos PLS2-DA e PLS1-DA mostra boas classificações e previsões para três grãos do tipo carioca: os cultivares Agro ANfc9, IPR-Sabiá e IPR-Andorinha. Em geral, a construção de três modelos binários PLS1-DA, tratando duas cultivares por vez, é ligeiramente melhor do que o modelo PLS2-DA obtido para as três classes simultaneamente. Os modelos PLS1-DA resultaram em melhor classificação e predição para as classes de cultivares Agro ANfc9 e IPR-Sabiá do que para a cultivar IPR-Andorinha. Além disso, ambos os modelos foram submetidos a testes de permutação que indicam que os modelos não estão sobreajustados. Embora os grãos estudados aqui tenham aparência semelhante, eles têm propriedades diferentes que são importantes para o consumidor e o produtor. Essa metodologia pode ser aplicada no controle de qualidade do feijão, verificando a autenticidade da cultivar. Por fim, essa técnica analítica é rápida, tem baixo custo e é ambientalmente correta, pois o preparo de amostra não gera resíduos.

REFERÊNCIAS

- Agro Norte Pesquisas e Sementes. (2020). *Semente de Feijão Carioca - ANfc 9*.
https://www.agronorte.com.br/Produtos/FeijaoCarioca/12-ANfc_9
- Antonelli, A., Cocchi, M., Fava, P., Foca, G., Franchini, G. C., Manzini, D., & Ulrici, A. (2004). Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-based classification algorithm. *Analytica Chimica Acta*, *515*, 3–13.
<https://doi.org/10.1016/j.aca.2004.01.005>
- Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods*, *5*(16), 3790–3798.
<https://doi.org/10.1039/c3ay40582f>
- Bassbasi, M., De Luca, M., Ioele, G., Oussama, A., & Ragno, G. (2014). Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data. *Journal of Food Composition and Analysis*, *33*(2), 210–215. <https://doi.org/10.1016/j.jfca.2013.11.010>
- Böhme, K., Calo-mata, P., Barros-Velázquez, J., & Ortea, I. (2019). Recent applications of omics-based technologies to main topics in food authentication. *Trends in Analytical Chemistry*, *110*, 221–232. <https://doi.org/10.1016/j.trac.2018.11.005>
- Byrne, L., Barker, J., Pennarun-Thomas, G., Diamond, D., & Edwards, S. (2000). Digital imaging as a detector for generic analytical measurements. *Trends in Analytical Chemistry*, *19*(8), 517–522.
- Chevallier, S., Bertrand, D., Kohler, A., & Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics*, *20*, 221–229.
<https://doi.org/10.1002/cem>

Cittolin Sementes. (2020). *IPR-Sabiá Características Agronômicas*.

Costa, N. P. da, Pereira, L. A. G., & Neto, J. de B. F. (1979). Método da peroxidase para identificação de cultivares de soja. *Conference Paper Present at I Congresso Brasileiro de Sementes*, 89–93.

Costa, G. B., Fernandes, D. D. S., Almeida, V. E., Araújo, T. S. P., Melo, J. P., Diniz, P. H. G. D., & Vêras, G. (2015). Digital image-based classification of biodiesel. *Talanta*, 139, 50–55. <https://doi.org/10.1016/j.talanta.2015.02.043>

Diniz, P. H. G. D. (2020). Chemometrics-assisted color histogram-based analytical systems. *Journal of Chemometrics*, 34(12), 1–27. <https://doi.org/10.1002/cem.3242>

Diniz, P. H. G. D., Dantas, H. V., Melo, K. D. T., Barbosa, M. F., Harding, D. P., Nascimento, E. C. L., Pistonesi, M. F., Band, B. S. F., & Araújo, M. C. U. (2012). Using a simple digital camera and SPA-LDA modeling to screen teas. *Analytical Methods*, 4, 2648–2652. <https://doi.org/10.1039/c2ay25481f>

Ferreira, B. C., Lima, S. F. de, Simon, C. A., Andrade, M. G. de O., Ávila, J. de, & Alvarez, R. de C. F. (2018). Effect of biostimulant and micronutrient on emergence, growth e quality of arabica coffee seedlings. *Coffee Science*, 13, 324–332.

Ferreira, C. A., Pinho, É. V. de R. Von, Paiva, P. D. de O., Pereira, G. S., & Diniz, R. P. (2011). Identificação de cultivares e certificação da pureza genética de gladiolo por meio de marcadores morfológicos. *Ciencia E Agrotecnologia*, 35, 692–700.

Fulgêncio, A. C. C., Araújo, V. P. T., Pereira, H. V., Botelho, B. G., & Sena, M. M. (2020). Development of a simple and rapid method for color determination in beers using digital images. *Food Analytical Methods*, 13, 303–312. <https://doi.org/10.1007/s12161-019-01634-0>

Gioia, T., Logozzo, G., Marzario, S., Zeuli, P. S., & Gepts, P. (2019). Evolution of SSR diversity from wild types to U.S. advanced cultivars in the Andean and Mesoamerican domestications of common bean (*Phaseolus vulgaris*). *PLoS ONE*, *14*, 1–21.

Grasel, F. S., Marcelo, M. C. A., & Ferrão, M. F. (2016). A non-destructive, rapid and inexpensive methodology based on digital images for the classification of natural tannin extracts. *RSC Advances*, *6*, 32358–32364. <https://doi.org/10.1039/c6ra00900j>

Grasel, F. S., Marcelo, M. C. A., & Ferrão, M. F. (2017). Development of an inexpensive, practical and non-destructive methodology based on digital images from a scanner for the classification of commercial tannins from *Acacia mearnsii*. *Analytical Methods*, *9*, 3977–3982. <https://doi.org/10.1039/c7ay00945c>

IDR-Paraná. (2013). *Cultivar de Feijão IPR Andorinha*.

Kelly, J. D. (2010). The Story of Bean Breeding. In *White paper prepared for BeanCAP & PBG Works on the topic of dry bean production and breeding research in the U.S.* (pp. 1–30).

Marcheafave, G. G., Tormena, C. D., Mattos, L. E., Liberatti, V. R., Ferrari, A. B. S., Rakocevic, M., Bruns, R. E., Scarminio, I. S., & Pauli, E. D. (2020). The main effects of elevated CO₂ and soil-water deficiency on ¹H NMR-based metabolic fingerprints of *Coffea arabica* beans by factorial and mixture design. *Science of the Total Environment*, *749*, 142350. <https://doi.org/10.1016/j.scitotenv.2020.142350>

Marini, F. (2013). Exploratory Data Analysis. In F. Marini (Ed.), *Chemometrics in Food Chemistry* (1st Ed, p. 513). Elsevier.

Mavromatis, A. G., Arvanitoyannis, I. S., Chatzitheodorou, V., Kaltsa, A., Patsiaoura, I., & Nakas, C. T. (2012). A comparative study among landraces of *Phaseolus vulgaris* L.

and *P. coccineus* L. based on molecular, physicochemical and sensory analysis for authenticity purposes. *Scientia Horticulturae*, *144*, 10–18. <https://doi.org/10.1016/j.scienta.2012.06.024>

Milanez, K. D. T. de M., & Pontes, M. J. C. (2015). Classification of extra virgin olive oil and verification of adulteration using digital images and discriminant analysis. *Analytical Methods*, *7*, 8839–8846. <https://doi.org/10.1039/c5ay01765c>

OECD. (2019). *Safety Assessment of Foods and Feeds Derived from Transgenic Crops* (Novel Food and Feed Safety, Ed.; Vol. 3). OECD Publishing. <https://doi.org/10.1787/f04f3c98-en>

Pedraño, M. R., & Coró, F. A. G. (1999). Análise sensorial e sua importância na pesquisa de alimentos. *Científica. Ciências Biológicas e Da Saúde*, *1*, 85–89.

Pereira, J. M., Leme, L. M., Perdoncini, M. R. F. G., Valderrama, P., & Março, P. H. (2018). Fast Discrimination of Milk Contaminated with *Salmonella* sp. Via Near-Infrared Spectroscopy. *Food Analytical Methods*, *11*(7), 1878–1885. <https://doi.org/10.1007/s12161-017-1090-0>

Pessoa, M. C. P. Y., Silva, A. de S., & Camargo, C. P. (2002). *Qualidade e Certificação de Produtos Agropecuários*.

Possobom, M. T. D. F., Ribeiro, N. D., Ribeiro, A. E. M., & Arns, F. D. (2015). Genetic control of the seed coat colour of Middle American and Andean bean seeds. *Genetica*, *143*, 45–54. <https://doi.org/10.1007/s10709-014-9811-4>

Santana, F. B. De, Gontijo, L. C., Mitsutake, H., Mazivila, S. J., Souza, L. M. De, & Borges Neto, W. (2016). Non-destructive fraud detection in rosehip oil by MIR

spectroscopy and chemometrics. *Food Chemistry*, 209, 228–233.
<https://doi.org/10.1016/j.foodchem.2016.04.051>

Santos, P. M., Simeone, M. L. F., Pimentel, M. A. G., & Sena, M. M. (2019). Non-destructive screening method for detecting the presence of insects in sorghum grains using near infrared spectroscopy and discriminant analysis. *Microchemical Journal*, 149, 104057. <https://doi.org/10.1016/j.microc.2019.104057>

Santos, P. M., Wentzell, P. D., & Pereira-Filho, E. R. (2012). Scanner Digital Images Combined with Color Parameters: A Case Study to Detect Adulterations in Liquid Cow's Milk. *Food Analytical Methods*, 5, 89–95. <https://doi.org/10.1007/s12161-011-9216-2>

Silochi, R. M. H. Q., Coelho, S. R. M., Bischoff, T. Z., Cassol, F. D. R., Padro, N. V. do, & Bassinello, P. Z. (2016). Nutritional technological characterization and secondary metabolites in stored carioca bean cultivars. *African Journal of Agricultural Research*, 11(24), 2102–2111. <https://doi.org/10.5897/AJAR2016.11013>

Silva, S. A. F., Silva, F. L. B., Ribas, A. F., Souza, S. G. H. de, & Santos, T. B. dos. (2020). Genome wide in silico analysis of SOD genes in common bean (*Phaseolus vulgaris* L.). *Journal of Crop Science and Biotechnology*, 23(3), 241–251. <https://doi.org/10.1007/s12892-020-00030-z>

Siqueira, S., Zaczuk, P., & Costa, S. (2016). Do enzymatic or non-enzymatic pathways drive the postharvest darkening phenomenon in carioca bean tegument? *LWT - Food Science and Technology*, 69, 593–600. <https://doi.org/10.1016/j.lwt.2016.01.079>

Snee, R. D. (1977). Validation of Regression Models: Methods and Examples Validation Examples of Regression Models: Methods and Examples. *Technometrics*, 19, 37–41. <https://doi.org/10.1080/00401706.1977.10489581>

Suhandy, D., & Yulia, M. (2017). Peaberry coffee discrimination using UV-visible spectroscopy combined with SIMCA and PLS-DA combined with SIMCA and PLS-DA. *International Journal of Food Properties*, 20(1), 331–339. <https://doi.org/10.1080/10942912.2017.1296861>

Szymanska, E., Saccenti, E., Smilde, A. K., & Westerhuis, J. A. (2012). Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomi*, 8(June), S3–S16. <https://doi.org/10.1007/s11306-011-0330-3>

Tormena, C. D., Marcheafave, G. G., Pauli, E. D., Bruns, R. E., & Scarminio, I. S. (2019). Potential biomonitoring of atmospheric carbon dioxide in *Coffea arabica* leaves using near-infrared spectroscopy and partial least squares discriminant analysis. *Environmental Science and Pollution Research*. <https://doi.org/10.1007/s11356-019-06163-1>

Valderrama, L., & Valderrama, P. (2016). Nondestructive identification of blue pen inks for documentoscopy purpose using iPhone and digital image analysis including an approach for interval confidence estimation in PLS-DA models validation. *Chemometrics and Intelligent Laboratory Systems*, 156, 188–195. <https://doi.org/10.1016/j.chemolab.2016.06.009>

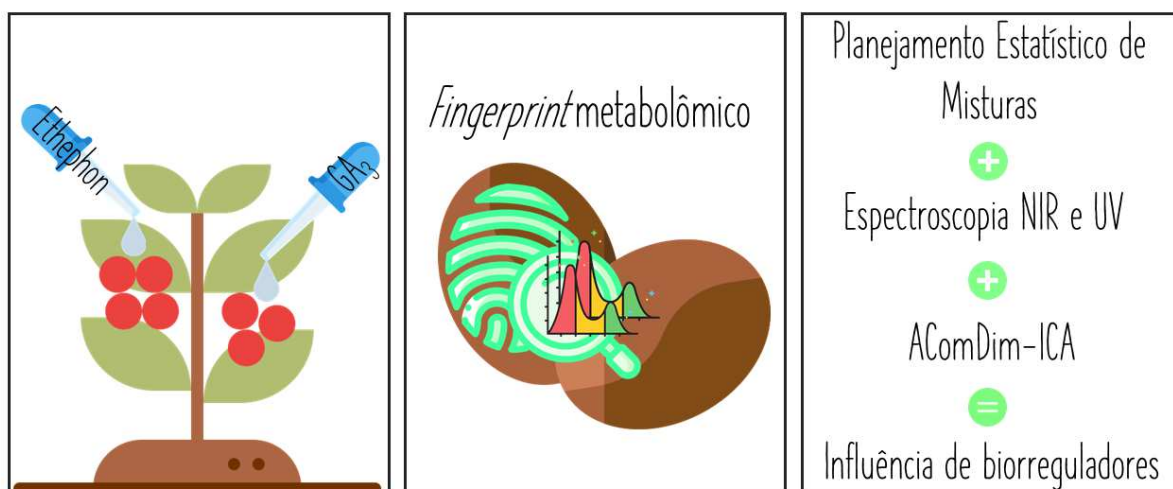
CAPÍTULO III

Aplicação exógena de biorreguladores em grãos de

Coffea arabica durante o amadurecimento:

Investigação de impressões digitais por design de

misturas usando UV, NIR e AComDim-ICA



RESUMO

Biorreguladores são substâncias utilizadas em diversos tipos de cultivos para otimizar sua quantidade e qualidade. Sua aplicação externa pode afetar o metaboloma da planta, desencadeando alterações metabólicas e conseqüentemente modificando a qualidade dos frutos. O presente estudo investigou as modificações metabômicas observadas nas impressões digitais espectrais de grãos de café cru dependentes da aplicação de biorreguladores (BIO, Giberelina e Ethephon) sobre o perfil vertical estratificado da planta (STR). A análise de componentes principais (PCA) foi utilizada no reconhecimento do solvente discriminante, obtido a partir do planejamento estatístico de mistura, para cada técnica aplicada. Ao analisar os espectros de infravermelho próximo (NIR) e UV-Visível (UV-Vis), os solventes **E** (etanol) e **ED** (etanol/diclorometano) foram o que apresentaram maior discriminação entre as amostras. Os espectros NIR e UV para estes extratos selecionados foram interpretados usando uma modificação da análise de variância-análise de dimensões comuns (AComDim), onde a etapa usual de PCA dentro do ComDim foi substituída por uma análise de componentes independentes (ICA), ou seja, AComDim-ICA. AComDim-ICA mostrou que o extrator **ED** foi o mais eficiente na discriminação do efeito biorregulador, enquanto o **E** foi o mais eficiente para a estratificação e as interações BIO×STR. A aplicação de AComDim-ICA também indicou que a giberelina permitiu a maior diferenciação em relação aos demais tratamentos BIO (Ethephon e controle), com possível influência nos compostos fenólicos. AComDim-ICA indicou que as bagas colhidas em estratos de até 1 m (altamente sombreadas) eram diferentes das colhidas em estratos mais altos (2 m e 3 m), relacionadas a metabólitos como cafeína, lipídios, açúcares e diterpenos. O metaboloma do estrato mais alto (3 m) tratado com Ethephon diferiu do controle para o mesmo estrato, em relação à cafeína. O uso desses métodos multivariados ampliou a análise metabômica das plantas, produzindo mais informações sobre o sistema.

1 INTRODUÇÃO

Biorreguladores são moléculas orgânicas de origem natural ou sintética que, em baixas concentrações, atuam no interior das células vegetais para estimular enzimas ou sistemas enzimáticos específicos e ajudam a regular o metabolismo vegetal, fortalecendo o sistema imunológico e reativando processos fisiológicos em diferentes estágios de desenvolvimento [1,2]. Uma única molécula de substância biorreguladora pode desencadear uma cascata de eventos metabólicos, resultando em alterações no desenvolvimento intracelular e crescimento externo. Além de melhorar a relação fonte-dreno e estimular a translocação de fotoassimilados em diferentes estágios de desenvolvimento, os biorreguladores também podem regular a produção de metabólitos secundários em plantas [3–5].

Nos últimos anos, o uso de biorreguladores na área agrônômica tem crescido, visando otimizar a quantidade e a qualidade de diversas culturas. O uso de substâncias biorreguladoras intervém em diversos processos fisiológicos e/ou morfológicos, incluindo capacidade fotossintética, germinação, crescimento vegetativo, formação efetiva de flores, frutos e, por fim, a produtividade das culturas, permitindo suas colheitas em períodos comercialmente desejáveis [1,5–7]. Seu uso na horticultura resultou na quebra da dormência das gemas, amadurecimento avançado dos frutos e floração estimulada, aumento do tamanho dos frutos e rendimento final da colheita [8–10].

Sendo uma das principais *commodities* brasileiras, o café em grão deve possuir uma qualidade excepcional para obter um alto valor agregado. A diferença de maturação entre os frutos de café (bagas) na mesma planta e no pomar, atrapalha a colheita e afeta a qualidade do produto [11]. Os grãos de café só devem ser colhidos quando atingirem a maturidade, numa fase em que já possuem as características

necessárias para gerar os aromas específicos do produto [12–14]. Portanto, as plantas de café em crescimento podem ser sensibilizadas internamente pela aplicação de biorreguladores para gerar produtividade agrônômica de melhor qualidade [1].

A aplicação externa de hormônios, como o Ethephon e os ácidos giberélicos, têm grande potencial para estimular o amadurecimento do café [7,15]. O Ethephon é o precursor sintético do etileno, o fitohormônio natural responsável por quebrar a dormência, estimular a floração, acelerar o amadurecimento das bagas [2,10,16] abscisão de órgãos, e senescência [17]. Os ácidos giberélicos (GA) são fitohormônios naturais, que estimulam a germinação de sementes, desencadeiam a transição do meristema para o crescimento da parte aérea, da fase juvenil para a folha adulta, do crescimento vegetativo para a floração, determinam a expressão sexual e o desenvolvimento do grão, além de interações com diferentes fatores ambientais, como a luz e fornecimento de água [18]. Embora a aplicação externa de Ethephon e GA₃ já tenha sido relatada em estudos com café [19] as alterações metabólicas das plantas que podem influenciar a qualidade final dos grãos, até onde sabemos, ainda não foram relatadas.

Considerando a complexidade química dos grãos de café (*Coffea arabica* L.), metodologias avançadas são necessárias para avaliar sua modificação química sob a ação de biorreguladores. A impressão digital metabolômica permite caracterizar dinamicamente qualquer influência, seja biótica, abiótica ou genética, em um organismo [20,21]. Devido à diversidade de constituintes das plantas, um processo que possa abranger todos os componentes metabolômicos dos grãos de café requer uma análise química complexa. Esse tipo de análise consiste em várias etapas multivariadas, por meio de amostragem de plantas definida por um planejamento fatorial [22], preparação da amostra por um planejamento estatístico de mistura [23,24], uso de

diferentes técnicas analíticas [25] e processamento de dados através de métodos quimiométricos apropriados [26,27].

A escolha adequada de solventes para o processo de extração na preparação da amostra pode ser uma etapa limitante para o isolamento de metabólitos da matriz global [28]. O planejamento estatístico de misturas já foi aplicado com sucesso em metabolômica para diversas matrizes vegetais, como grãos de café verde e torrado [29–31], *Mikania laevigata* [32], *Erythrina speciosa* [33,34], *Rollinia mucosa* [28], *Bauhinia* spp. [33], *Ilex paraguariensis* [35,36], entre outros. Esta metodologia tem se mostrado eficiente na ampliação da análise metabolômica, pois diversos solventes com diferentes propriedades físico-químicas são escolhidos para maximizar o alcance das impressões digitais espectroscópicas e cromatográficas adquiridas [37].

Técnicas espectroscópicas, tais como a espectroscopia ultravioleta (UV) e infravermelho próximo (NIR), têm sido usadas para identificar e caracterizar alguns dos metabólitos contidos nos grãos de café [11,22,24,30,38,39]. Devido ao baixo custo do equipamento e preparação de amostras relativamente simples, essas técnicas podem ser usadas como ferramenta de triagem, para discriminar vários padrões relacionados a metabólitos que absorvem luz em diferentes comprimentos de onda [40–43]. Essas técnicas analíticas podem ser associadas a abordagens quimiométricas adequadas para otimizar a extração de informações simultaneamente e extrair informações para a identificação de padrões ou características de interesse [44–46].

Vários estudos mostraram que quaisquer variações durante o cultivo, como clima, fornecimento de água, condições de luz solar e aumento da concentração de CO₂, podem afetar o metabolismo primário e, conseqüentemente, afetar o metabolismo secundário. Assim, supõe-se que qualquer modificação externa, como a aplicação externa de biorreguladores, pode impactar o metaboloma da planta e assim pode ser diagnosticada

através de perfis químicos do material vegetal. Então, entender como as mudanças metabólicas impactam o metaboloma da planta pode ajudar a melhorar posteriormente a qualidade final dos produtos. Nesse sentido, levantamos a hipótese de que grãos de café com aplicação externa de biorreguladores para padronização da maturação podem induzir alterações metabólicas que influenciam diretamente nos metabólitos que levam a uma bebida de qualidade. Assim, o objetivo deste trabalho foi investigar possíveis modificações metabolômicas refletidas nos espectros de grãos de café resultantes da aplicação de um biorregulador (Ethephon ou GA₃) em bagas verdes antes de seu amadurecimento final sobre o perfil vertical estratificado da planta, recebendo diferentes quantidades e qualidades de luz. Para fazer isso, os espectros foram analisados usando uma modificação do método da análise de variância – análise de dimensão comum (AComDim), onde a etapa usual de análise de componentes principais (PCA) dentro do ComDim é substituída por uma análise de componentes independentes (ICA), ou seja, AComDim-ICA. Isso pode facilitar a identificação de variações nos dados resultantes do efeito dos fatores do delineamento experimental, biorreguladores e estratificação, bem como sua interação.

2 MATERIAIS E MÉTODOS

2.1 AMOSTRAS

O experimento foi realizado na Embrapa Meio Ambiente, Jaguariúna-SP, Brasil (22°43'S, 47°01'O, 570 m de altitude). As plantas de café tinham 8 anos, irrigadas, com cerca de 2,7-3 m de altura. Quando todas as bagas estavam expandidas e verdes, um mês e meio antes da colheita, as plantas foram pulverizadas, de cima para baixo, com GA₃ ou Ethephon para promover um amadurecimento mais uniforme, enquanto as plantas controle foram pulverizadas com a mesma quantidade de água destilada.

Para a preparação de GA₃, foi feita uma solução estoque utilizando 1 g de GA₃ (C₁₉H₂₂O₆, Sigma-Aldrich) dissolvido em 10 mL de etanol puro e depois diluído em 1 L de água destilada, para obter uma concentração de 1000 ppm. Para preparar a solução de aplicação, 25 mL da solução estoque de GA₃ foram diluídos em 500 mL de água destilada, para obter uma concentração de GA₃ de 50 ppm com uma gota de Tween™ 20, e foram aplicados 125 mL desta solução por planta. A preparação do Ethephon também utilizou uma solução estoque composta por 0,1 g (100 mg) de Ethephon puro (ácido 2-cloroetilfosfônico, 96%, Sigma-Aldrich) diluído em 100 mL de água destilada para obter uma concentração de 1000 ppm. Para a solução de aplicação, 25 mL da solução estoque de Ethephon foram diluídos em 500 mL de água destilada para obter uma concentração de 50 ppm com uma gota de Tween™ 20, e foram aplicados 125 mL por planta. As aplicações de GA₃, Ethephon e água destilada foram realizadas em 12 de abril de 2018 e sempre feitas até que o líquido pingasse das bagas, mesmo que também fossem atingidos folhas e galhos.

Os grãos de café arábica foram coletados em três estratos do perfil vertical das plantas: até 100 cm (1 m), de 101 a 200 cm (2 m) e acima de 201 cm (3 m). A colheita das bagas foi feita de 23 a 25 de maio de 2018. As bagas foram secas em corrente de ar, à sombra. Os grãos secos (um a dois grãos estão contidos em uma baga) foram separados da casca e impurezas, resultando em grãos crus (verdes) representativos, dos quais os defeituosos foram excluídos antes da moagem para análise.

Para avaliar se os fatores de biorregulação e estratificação ou sua interação tiveram um efeito estatisticamente significativo na impressão digital metabolômica de grão cru *C. arabica*, a análise metabólica foi realizada seguindo um planejamento fatorial completo de três níveis com dois fatores (3^2) - (Tabela 1).

Tabela 1. Delineamento fatorial com dois níveis e três fatores (3^2) para análise da aplicação de biorreguladores e efeitos da estratificação nas impressões digitais metabolômicas de grãos de *Coffea arabica*.

Fatores	-1	0	+1
1. Bioregulador	Ethephon	Controle	GA ₃
2. Estratificação	1 metro	2 metros	3 metros
	Nível do Fator		
Experimento	1	2	
1	-1	-1	
2	-1	0	
3	-1	+1	
4	0	-1	
5	0	0	
6	0	+1	
7	+1	-1	
8	+1	0	
9	+1	+1	

2.2 EXTRAÇÃO

A impressão digital pode ser entendida como uma representação do metabolismo da planta em um momento exato refletindo as concentrações de seus metabólitos observáveis [20]. Uma de suas limitações é o processo extrativo, que deve refletir o metaboloma total da planta [25]. Para otimizar este processo extrativo para metabólitos de *C. arabica*, foi utilizado um planejamento estatístico de mistura composto por quatro solventes. A escolha dos solventes foi baseada no triângulo de seletividade de Snyder [47]. Neste caso específico, os solventes utilizados foram etanol (**E**), diclorometano (**D**), acetona (**A**) e hexano (**H**). Diclorometano, etanol e hexano foram obtidos da Anidrol (Diadema, Brasil) e acetona da Vetec (Rio de Janeiro, Brasil).

As extrações para esta análise exploratória foram realizadas em grãos de *C. arabica* crus, de bagas coletadas de diferentes estratos vegetais tratados com Ethephon e GA₃ mais os controles. O ponto central do planejamento foi realizado em triplicata, resultando em 17 extratos (Tabela 2) do planejamento de misturas (ou seja, totalizando 51 extratos para cada biorregulador ou condição de controle, já que cada ponto do planejamento foi realizado em triplicata). Para tanto, foram utilizados 6,0 g de grãos crus moídos, aos quais foram adicionados 60,0 mL da mistura de solventes. Cada mistura foi colocada em banho ultrassônico por 60 minutos a 20 °C e filtrada em papel filtro para separar a solução dos grãos de café. Esses mesmos grãos crus foram novamente submetidos à extração exaustiva por mais seis vezes, totalizando 420 mL para cada extração. Os extratos foram concentrados e secos em uma corrente de ar seguida de liofilização. Para as análises espectrais NIR e UV-Vis, 10 mg de cada amostra liofilizada foram solubilizados em seu respectivo solvente. Após a análise exploratória dos solventes e a escolha do solvente mais discriminante, novas extrações foram realizadas com os

solventes selecionados para buscar as diferenças metabólicas entre os grãos dos três estratos dos cafeeiros submetidos aos diferentes tratamentos.

Tabela 2. Proporções do planejamento estatístico de misturas para os quatro componentes: etanol, diclorometano, acetona e hexano.

Ponto do Planejamento	Notação	Solventes			
		Etanol	Diclorometano	Acetona	Hexano
1	E	1	0	0	0
2	D	0	1	0	0
3	A	0	0	1	0
4	H	0	0	0	1
5	ED	1/2	1/2	0	0
6	EA	1/2	0	1/2	0
7	EH	1/2	0	0	1/2
8	DA	0	1/2	1/2	0
9	DH	0	1/2	0	1/2
10	HA	0	0	1/2	1/2
11	EDA	1/3	1/3	1/3	0
12	EDH	1/3	1/3	0	1/3
13	EHA	1/3	0	1/3	1/3
14	DAH	0	1/3	1/3	1/3
15	EDAH1	1/4	1/4	1/4	1/4
16	EDAH2	1/4	1/4	1/4	1/4
17	EDAH3	1/4	1/4	1/4	1/4

2.3 UV-VIS

Os espectros de absorção UV-Visível foram registrados usando um espectrofotômetro UV-Vis Evolution 60S (Thermo Fisher Scientific, Waltham, Massachusetts, EUA), em cubetas de quartzo com 1 cm de caminho óptico, acoplado ao software VISION Lite (Thermo Fisher Scientific, Waltham, Massachusetts, EUA). Os espectros foram adquiridos entre 190 e 800 nm, a 22 ± 1 °C. As aquisições foram feitas com 20 µL do extrato solubilizado e diluído em 2 mL de seu respectivo solvente extrator. Os conjuntos de dados foram coletados em unidades de absorbância e salvos como arquivos de dados ASCII.

2.4 INFRAVERMELHO PRÓXIMO (NIR)

Os espectros de transmitância NIR foram adquiridos entre 900 e 1700 nm com resolução de 1 nm com o DLP NIRscan Nano (Texas Instruments Inc., Dallas-TX, EUA) com de 1,0 mL de extrato solubilizado na cubeta de quartzo em temperatura ambiente de 22 ± 1 °C. Cada espectro representa uma média de 99 varreduras. Os espectros foram adquiridos em triplicata e o valor médio foi utilizado.

2.5 ANÁLISE EXPLORATÓRIA PARA ESCOLHA DO SOLVENTE DE EXTRAÇÃO

A análise de componentes principais (PCA) é uma ferramenta exploratória usada para reduzir a dimensionalidade de grandes conjuntos de dados, transformando as variáveis originais em novas variáveis chamadas de componentes principais (PCs) que não são correlacionadas e são ordenadas no sentido de que o primeiro componente apresenta a maior parte da variação contida em todas as variáveis originais [48,49]. As PCs ajudam a identificar possíveis formações de padrões amostrais devido a diferenças e semelhanças na forma como as variáveis influenciam a dispersão das amostras [50,51].

A PCA foi usada para encontrar o solvente mais promissor, com base nos resultados de um planejamento estatístico de mistura de solventes, para separar os espectros de grãos de cafeeiros que foram tratados com Ethephon ou GA₃ do grupo controle (água destilada). A PCA foi realizada separadamente nos dados de NIR e UV. Os espectros NIR de grãos de café com e sem aplicação de biorregulador foram dispostos em uma matriz de 153 linhas e 228 colunas, enquanto a matriz UV-Vis continha 153 linhas e 611 colunas. As linhas representaram os diferentes solventes (51 extratos para cada biorregulador ou condição de controle) utilizados no processo extrativo, de acordo com o planejamento estatístico de mistura (Tabela 2), enquanto os dados nas colunas

foram os diferentes comprimentos de onda. Antes de aplicar a PCA às matrizes de dados, vários métodos de pré-processamento foram testados. Para o NIR, os melhores resultados foram alcançados pela suavização de Savitzky-Golay com janelas de 9 pontos, segunda ordem polinomial e primeira derivada. Os espectros UV-Vis foram tratados pelos mesmos métodos, mas com uma janela de 15 pontos. Ambos os dados NIR e UV-Vis foram centrados na média.

2.6 ANÁLISE ACOMDIM-ICA

Para determinar os efeitos dos biorreguladores e da estratificação na qualidade metabólica do grão de café, as variações da intensidade do sinal nas duas regiões espectrais foram comparadas com outras fontes de variação (ruído ou erro experimental). Para isso, aplicou-se o método AComDim para verificar se as variações espectrais devido a uma mudança no nível do fator eram maiores que a variabilidade residual. AComDim baseia-se na mesma ideia da análise de variância-componentes principais [52], na qual a matriz de dados experimental é decomposta em uma série de matrizes sucessivas contendo, para cada fator e interação, os vetores médios de cada nível. A matriz residual (calculada após a subtração sucessiva de cada matriz média) é então adicionada de volta a cada matriz de médias. Uma PCA é então aplicada a cada matriz de médias mais resíduos. Para tanto, a PCA é utilizada para avaliar a significância de cada fator comparando sua variância com a do erro residual [53]. Se os escores das amostras são distribuídos ao longo da PC1 de forma que reflita os níveis dos fatores correspondentes e de forma aleatória ao longo das outras PCs, então o efeito desse fator é considerado como tendo um efeito significativo sobre os valores na matriz de dados original. Com o AComDim, as PCs separados e individuais são substituídos por uma

única análise simultânea de todas as matrizes médias mais resíduos em conjunto, bem como a matriz de resíduos, usando o ComDim [54].

ComDim é uma implementação particular de componentes comuns e análise de pesos específicos (do inglês, *common components and specific weights analysis*, CCSWA), que é descrita por Meyners [55]. O método calcula iterativamente cada componente comum (CC) sucessiva que representa a direção de dispersão máxima das amostras em um espaço comum a todas as tabelas de dados. Cada tabela tem um peso específico, também chamado de saliência [56,57]. As saliências, no caso do AComDim, indicam qual fator ou interação mais contribui para cada CC [54]. Uma vez que os resíduos estão presentes em todas as matrizes; os blocos que contribuem significativamente para o CC1 (aqueles com altas saliências) contêm principalmente ruído, enquanto os blocos que contribuem menos contêm outras fontes de variabilidade além do ruído [54]. Para estimar a significância de um bloco, um teste de Fisher (teste F) [58] pode ser aplicado aos valores F (F_i) calculados como segue (Eq. (1)).

$$F_i = \frac{\lambda_{res}}{\lambda_i} \quad \text{Eq. (1)}$$

onde λ_{res} é a saliência do bloco residual em CC1, e λ_i é a saliência do i -ésimo bloco em CC1.

Os blocos para os quais F_i é maior que o valor crítico (F_c) da tabela de Fisher são considerados como estando relacionados a fatores ou interações influentes [59]. O teste F é realizado selecionando $n - 1$ graus de liberdade, onde n é o número de linhas em cada matriz. O nível de significância é igual a 0,05. No entanto, às vezes o bloco de resíduos é o principal contribuinte para vários CCs, não apenas CC1, refletindo várias fontes ortogonais de variabilidade residual. Nesse caso, divide-se a soma das saliências residuais em todos esses CCs pela soma das saliências de cada bloco nos mesmos CCs [60]. No presente estudo, as matrizes dos extratos obtidos tanto com etanol

puro (**E**) quanto com a mistura de etanol e diclorometano (**ED**) foram analisadas simultaneamente.

Uma implementação mais recente do ComDim [61] torna possível substituir a decomposição PCA do conjunto de dados multibloco por uma análise de componentes independentes (ICA). O ICA é um método de separação de fonte cega que visa extrair os sinais subjacentes puros de um conjunto de dados onde eles são misturados em proporções desconhecidas [62,63]. Esses sinais de origem são semelhantes aos *loadings* da PCA, mas podem estar mais diretamente relacionados a fenômenos específicos que impactam as variáveis, enquanto as proporções correspondentes são semelhantes aos escores de PCA, mas podem estar diretamente relacionadas à intensidade de cada fenômeno [64].

O AComDim-ICA foi realizado separadamente nos dados de NIR e UV-Vis, em que cada matriz continha 297 espectros de amostra, 33 para cada uma das nove combinações agronômicas (Tabela 1), obtidas a partir de novas extrações com **E** e **ED**. Para a análise do NIR, os dados espectrais foram segmentados em quatro blocos: espectros de grãos de café de extratos etanólicos na faixa entre 905 a 1615 nm (**E1**) e 1652 a 1698 nm (**E2**), e extratos de mistura de etanol e diclorometano na faixa entre 905 a 1615 nm (**ED1**) e 1652 a 1698 nm (**ED2**). O conjunto de dados espectrais UV-Vis foi organizado em dois blocos: extratos de grãos de café obtidos por **E** na faixa entre 205 e 400 nm e **ED** de 225 a 400 nm. Essas faixas espectrais deram os melhores resultados e menos ruído devido a solventes. Antes do AComDim-ICA, os espectros NIR foram pré-processados por primeira derivada e espectros UV-Vis por segunda derivada.

2.7 ANÁLISE ESTATÍSTICA E SOFTWARE

Todos os cálculos foram realizados usando o software Matlab 2016® (R2016b). O procedimento AComDim-ICA foi adaptado da função ComDim na caixa de ferramentas gratuita SAISIR [65] e do algoritmo JADE para ICA [66]. As funções estão disponíveis em: <https://github.com/DNRUtlledge/AComDim>.

3 RESULTADOS

3.1 ANÁLISE EXPLORATÓRIA

Já foi demonstrado que modificações externas podem alterar metabólitos presentes no grão de café, alterando a qualidade final da bebida [11,67]. A primeira etapa desta pesquisa propriamente dita foi constituída por análises exploratórias dos efeitos do solvente na extração metabolômica e das variações metabólicas causadas pelo amadurecimento dos frutos sob efeito de aplicação hormonal externa. Para isso, a PCA foi aplicada aos conjuntos de dados dos espectros de transmitância NIR e UV-Vis.

As impressões digitais do NIR foram usadas para discriminar metabolicamente os grãos das bagas induzidas a amadurecer por aplicação de hormônio externo (biorreguladores) daquelas que amadureceram em condições naturais (controle) e para determinar qual dos solventes utilizados permitiria essa discriminação. As duas primeiras PCs para os dados NIR corresponderam a 84,25% da variância total explicada dos dados (Fig. 1). Elas sugeriram que o etanol (E) e a mistura binária etanol-diclorometano (ED) apresentaram o melhor potencial como extratores para a separação metabolômica de grãos de café, quando biorreguladores foram aplicados.

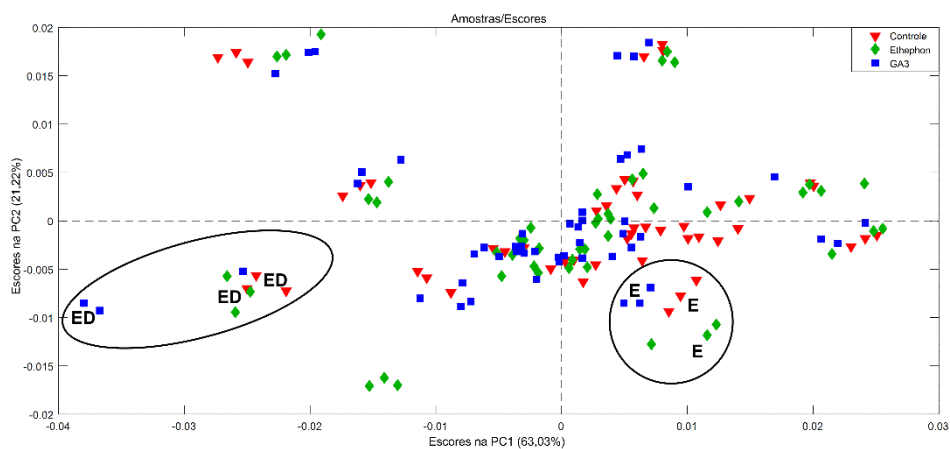


Figura 1. Gráficos dos escores na primeira e segunda componentes principais (PCs) das impressões digitais de transmitância NIR de grãos de *Coffea arabica* adquiridos a partir

dos extratos do planejamento estatístico de mistura. Os grãos foram provenientes de plantas tratadas com água destilada - Controle (▼), Ethephon (◆) ou GA₃ (■).

Os escores das duas primeiras componentes na análise de impressão digital UV-Vis (Fig. 2) corresponderam a 92,93% da variância total dos dados. Assim como no caso da transmitância NIR (Fig. 1), o etanol (E) e a mistura binária de etanol e diclorometano (ED) mostraram a melhor discriminação metabolômica entre grãos de café de plantas tratadas com biorreguladores daquelas onde as bagas amadureceram naturalmente (Fig. 2). Portanto, a análise exploratória para a escolha do melhor solvente discriminante mostrou que, para os grãos de café, E e ED puros se destacaram tanto para as técnicas NIR quanto UV. Esses resultados corroboram com artigos já publicados em que estes mesmos solventes são utilizados para discriminações de metabólitos de grãos de café a partir de técnicas de NIR e UV [22,30].

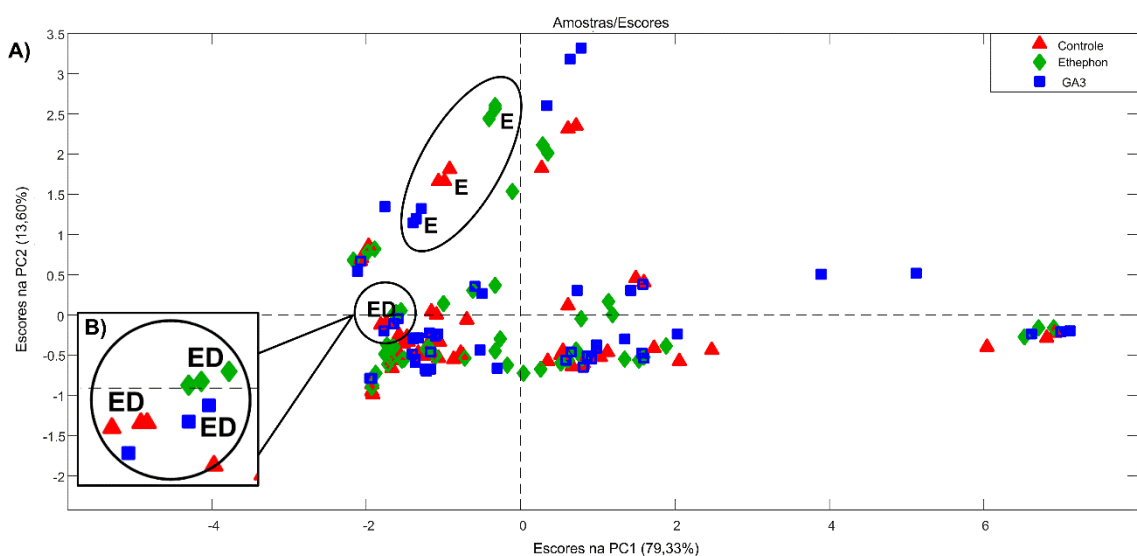


Figura 2. A) Gráficos dos escores da primeira e segunda componentes principais (PC) das impressões digitais UV-Vis de grãos de *Coffea arabica* adquiridos a partir dos extratos do planejamento estatístico de mistura. B) A aproximação dos gráficos dos escores mostra mais claramente as amostras ED. Os grãos eram de plantas tratadas com água destilada - Controle (▼), Ethephon (◆) ou GA₃ (■).

3.2 AComDIM-ICA

Após a triagem inicial apresentada na seção 3.1, novas extrações foram realizadas com os solventes **E** e **ED**, porém utilizando um número maior de amostras para avaliar tanto a aplicação dos biorreguladores quanto a estratificação vertical da planta. A estratificação foi incluída na análise exploratória para determinar se as variações ambientais externas inerentes ao perfil vertical da planta, principalmente relacionadas às variações de luz e temperatura [67,68], interfeririam no efeito das aplicações de biorreguladores. Assim, foi realizada uma análise multibloco global em todas as matrizes correspondentes aos dois fatores (biorreguladores e estratificação), bem como suas interações e os resíduos, utilizando AComDim-ICA. O objetivo do AComDim-ICA foi avaliar possíveis alterações espectrais devido aos efeitos biorreguladores e de estratificação. Para os dois domínios espectrais, 9 CCs foram computados.

3.2.1 ANÁLISE MULTIBLOCO NIR

Para o estudo do efeito de biorreguladores e estratificação nos metabólitos do grão de café usando a região espectral NIR, o AComDim-ICA aplicado a **E1**, **E2**, **ED1** e **ED2** forneceu os valores de saliência desses blocos em cada CC (Fig. 3). Como esperado, CC1 caracterizou o ruído residual para o **ED1** (barra verde), já que a maior saliência para CC1 foi para o bloco de resíduos. Além disso, a variabilidade residual no CC1 foi maior do que nos outros oito CCs. Situação semelhante foi encontrada para CC7 e CC8 (ambos caracterizando componentes do ruído residual para **ED2** (barra roxa), CC2 e CC6 para ruído residual **E1** (barra vermelha) e CC9 para ruído residual **E2** (barra azul). Em CC3, os biorreguladores (bloco 1) tiveram valores de saliência muito maiores que os demais fatores para **ED1** e **ED2**; no CC4, o fator de estratificação teve a maior saliência para **E2**, e CC5 a interação entre biorreguladores e estratificação teve as maiores saliências para **E2** e **ED2**.

Os valores F dos blocos foram calculados com base em CCs com altas saliências para os resíduos, para cada faixa espectral e bloco (**E1**, **E2**, **ED1** e **ED2**) – (Fig. 4). Esses valores F foram comparados com o valor crítico da tabela de Fisher de acordo com o nível $\alpha = 0,05$, promovendo um valor crítico de 1,21 (296 graus de liberdade). Os valores F para as razões do conjunto de dados espectrais NIR para o fator biorregulador (BIO: bloco 1) foram 2,15 e 2,31 para os conjuntos de dados **ED1** e **ED2**, respectivamente (Fig. 4). Esses valores foram superiores ao valor crítico, indicando que as variações espectrais para esses extratos foram significativas em relação aos resíduos. Essas duas matrizes apresentaram as maiores saliências em CC3. Os conjuntos de dados do extrato **E2** foram significativamente diferentes dos resíduos, com valores F iguais a 2,96 para o fator de estratificação (STR: bloco 2) e 1,97 para a interação (BIO×STR: bloco 3). A matriz **E2** apresentou os maiores valores de saliência em CC4 para o fator de estratificação, e os maiores valores de saliência em CC5 para a interação BIO×STR.

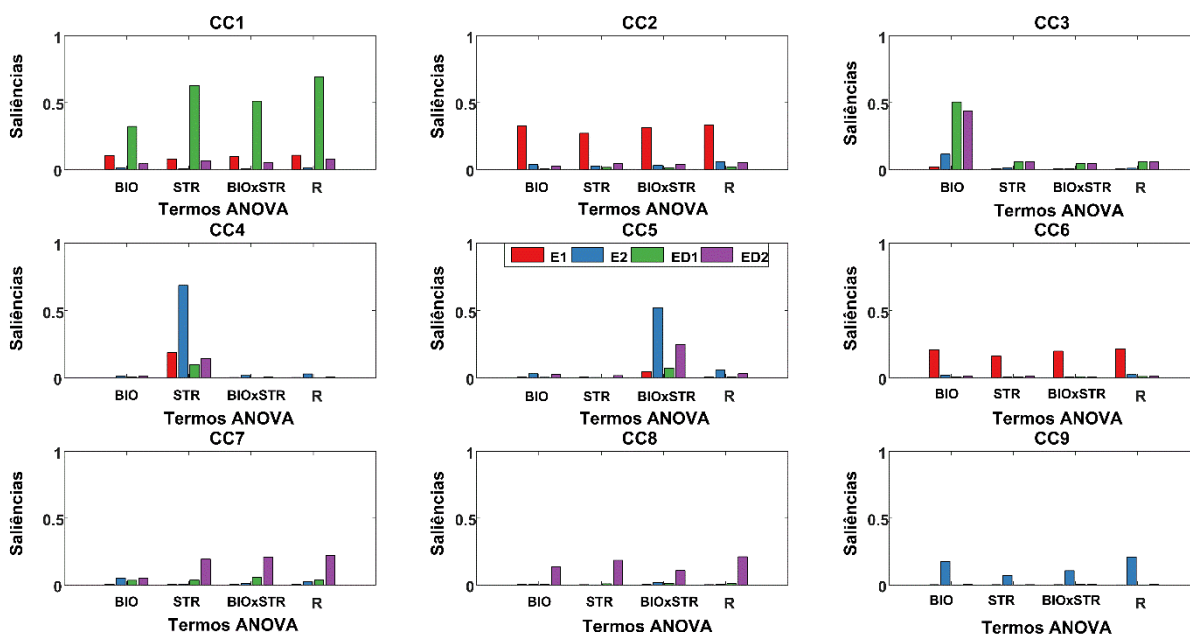


Figura 3. Saliências dos blocos decompostos ANOVA, para as Componentes Comuns (CC) 1 a 9, em que BIO representa o fator biorregulador, STR o fator de estratificação, BIO×STR a interação entre biorreguladores e estratificação e, finalmente, R a matriz

residual. As barras foram **E1** (etanol/905 a 1615 nm - vermelho), **E2** (etanol/1652 a 1698 nm - azul), **ED1** (etanol e diclorometano/905 a 1615 nm - verde), **ED2** (etanol e diclorometano/1652 a 1698 nm - roxo).

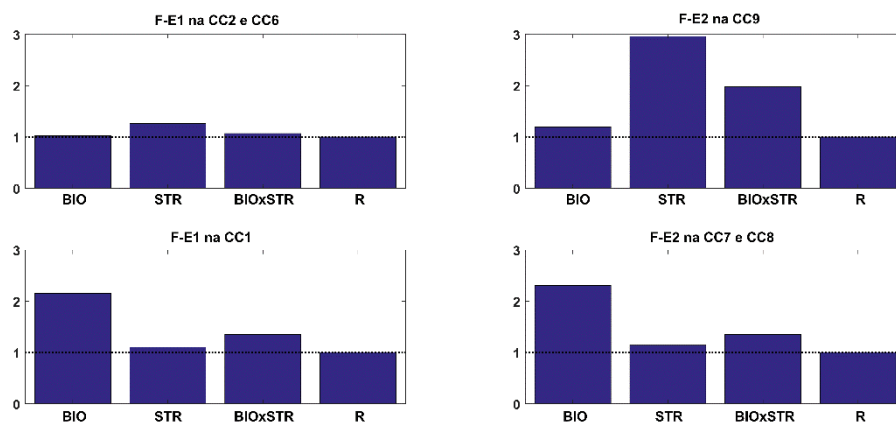


Figura 4. Valores F baseados nos Componentes Comuns (CC) 1 a 9, com altas saliências para resíduos calculados a partir do AComDim-ICA de impressões digitais espectrais NIR para grãos de café originados de plantas controle ou plantas tratadas com biorreguladores e colhidas em diferentes estratos ao longo do perfil vertical da planta. BIO representa o fator biorregulador, STR o fator de estratificação, BIO×STR a interação entre biorreguladores e estratificação e R a matriz residual.

Os CC3, 4 e 5 foram relacionados aos fatores BIO, STR e BIO×STR, respectivamente, e estes deram origem a agrupamentos das amostras com base nos níveis desses três casos nos gráficos dos escores correspondentes (Fig. 5), enquanto os *loadings* associados são dados na Fig. 6. O gráfico dos escores CC3×CC1 indicou amostras que foram separadas pelo efeito biorregulador (Fig. 5a). As amostras que receberam aplicação de GA₃ foram projetadas no lado positivo do CC3. Com base nos *loadings* positivos em CC3 (Fig. 6a), esses extratos de café podem apresentar metabólitos característicos de polifenóis, carboidratos e cafeína/metilxantinas devido à presença de bandas em torno de 1158 nm, 1455 e 1685 nm, respectivamente [38,69–71]. As amostras controle

apresentaram escores mais negativos em CC3 (Fig. 5a), o que pode estar associado à extração de açúcares (1409 nm), cafeína/metilxantinas (1691 e 1670 nm) e ácido clorogênico (1652 nm) - (Fig. 6a) - [38,69,72,73]. Os extratos responsáveis por esta separação metabólica foram obtidos utilizando um solvente extrator contendo etanol e diclorometano (ED).

O gráfico dos escores CC4×CC1 (Fig. 5b) mostrou a separação dos extratos de café de acordo com a posição do perfil vertical do fruto. Grãos de café de frutos colhidos no estrato mais baixo (1 m acima do solo), foram projetados mais no lado positivo do CC4. As amostras dos estratos vegetais superiores (2 e 3 m) foram bem separadas no lado negativo de CC4. Grãos de café colhidos até 1 m da altura da planta podem apresentar concentrações relativas elevadas de cafeína, indicadas por bandas em torno de 1682 e 1695 nm, de acordo com os valores de *loadings* de E2, em comparação com os outros dois estratos (Fig. 6b) - [38,72,74].

A partir dos escores de CC5×CC1 relacionados à interação BIO×STR (Fig. 5c) pode-se notar uma diferença nas respostas metabólicas entre os grãos que receberam Ethephon nos três estratos vegetais, quando comparados às amostras do grupo controle. Em outras palavras, os grãos de café coletados do estrato mais alto (3 m) das plantas tratadas com Ethephon (Fig. 5c-3) tiveram valores de *loadings* mais positivos no CC5 do que no controle-3 m (Fig. 5c-6), bem como GA₃-3 m (Fig. 5c-9) (experimentos 3, 6 e 9 da Tabela 1, respectivamente). De acordo com os valores dos escores (E2), as bandas em torno de 1679 e 1695 nm foram relacionadas ao Ethephon-3 m (Fig. 5c-3), provavelmente devido à elevada quantidade de cafeína (Fig. 6c) - (Pizarro et al., 2007). A comparação semelhante pode ser feita entre Ethephon-1 m (Fig. 5c-1) e 2 m (Fig. 5c-2) (experimentos 1 e 2 da Tabela 1, respectivamente), que estavam no lado negativo de CC5, com o controle-1 m (Fig. 5c-4) e 2 m (Fig. 5c-5) (experimentos 4 e 5 da Tabela 1,

respectivamente) e o GA₃-1 m (Fig. 5c-7) e 2 m (Fig. 5c-8) (experimentos 7 e 8 da Tabela 1, respectivamente), que estavam no lado positivo deste CC (Fig. 5c). De acordo com os valores de *loadings* (E2) (Fig. 6c), o grupo controle e os frutos tratados com GA₃ colhidas nos estratos mais baixos (1 m) e no meio (2 m) tiveram as maiores concentrações relativas de cafeína relacionadas às bandas em torno de 1679 e 1695 nm [72], em comparação com frutos tratados com Ethephon dos mesmos estratos.

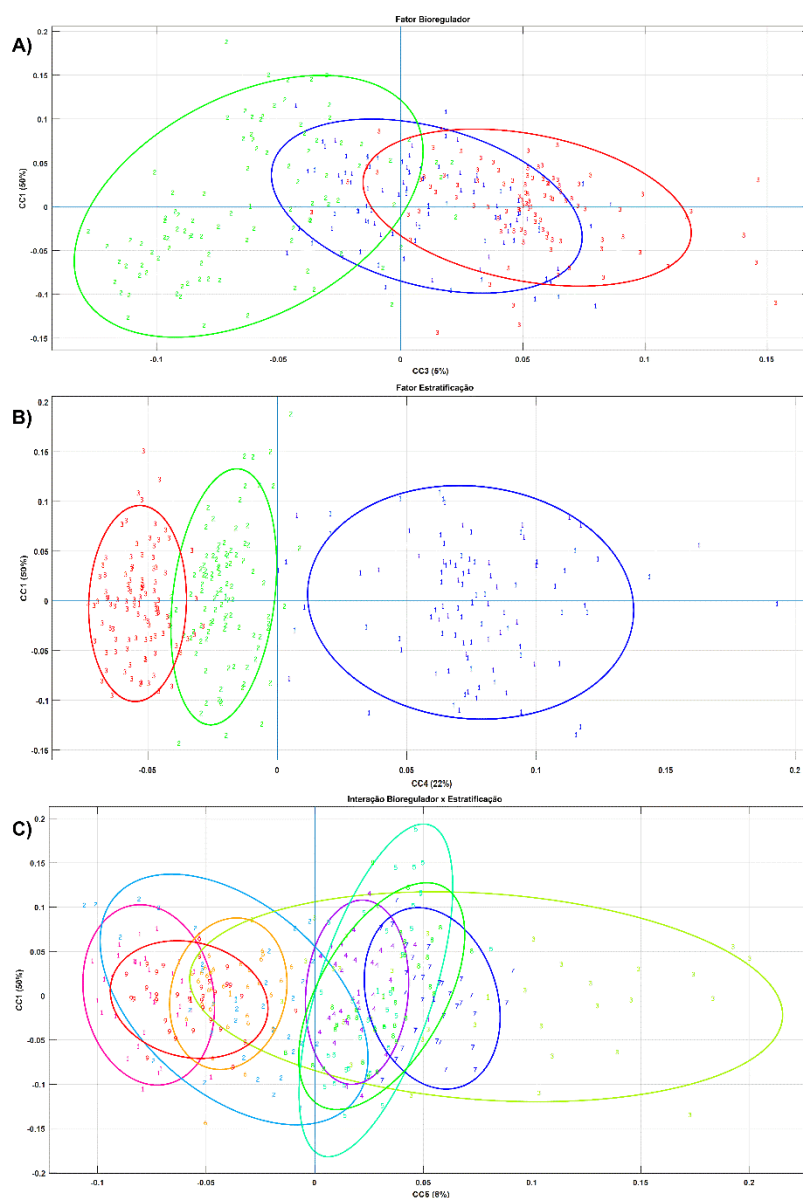


Figura 5. Gráficos dos escores AComDim-ICA das impressões digitais espectrais NIR de grãos de café originados de plantas tratadas com biorreguladores e colhidas nos três

estratos verticais de plantas: A) Gráfico dos escores $CC1 \times CC3$ para fator biorregulador - (1) Ethephon, (2) Controle e (3) GA_3 ; B) Gráfico dos escores $CC1 \times CC4$ para fator de estratificação - (1) o estrato mais baixo - 1 m, (2) o estrato médio - 2 m e (3) o estrato mais alto - 3 m; C) Gráfico dos escores $CC1 \times CC5$ para a interação “biorregulador \times estratificação” - (1) Ethephon-1 m, (2) Ethephon-2 m, (3) Ethephon-3 m, (4) Controle-1 m, (5) Controle-2 m, (6) Controle-3 m, (7) GA_3 -1 m, (8) GA_3 -2 m e (9) GA_3 -3 m.

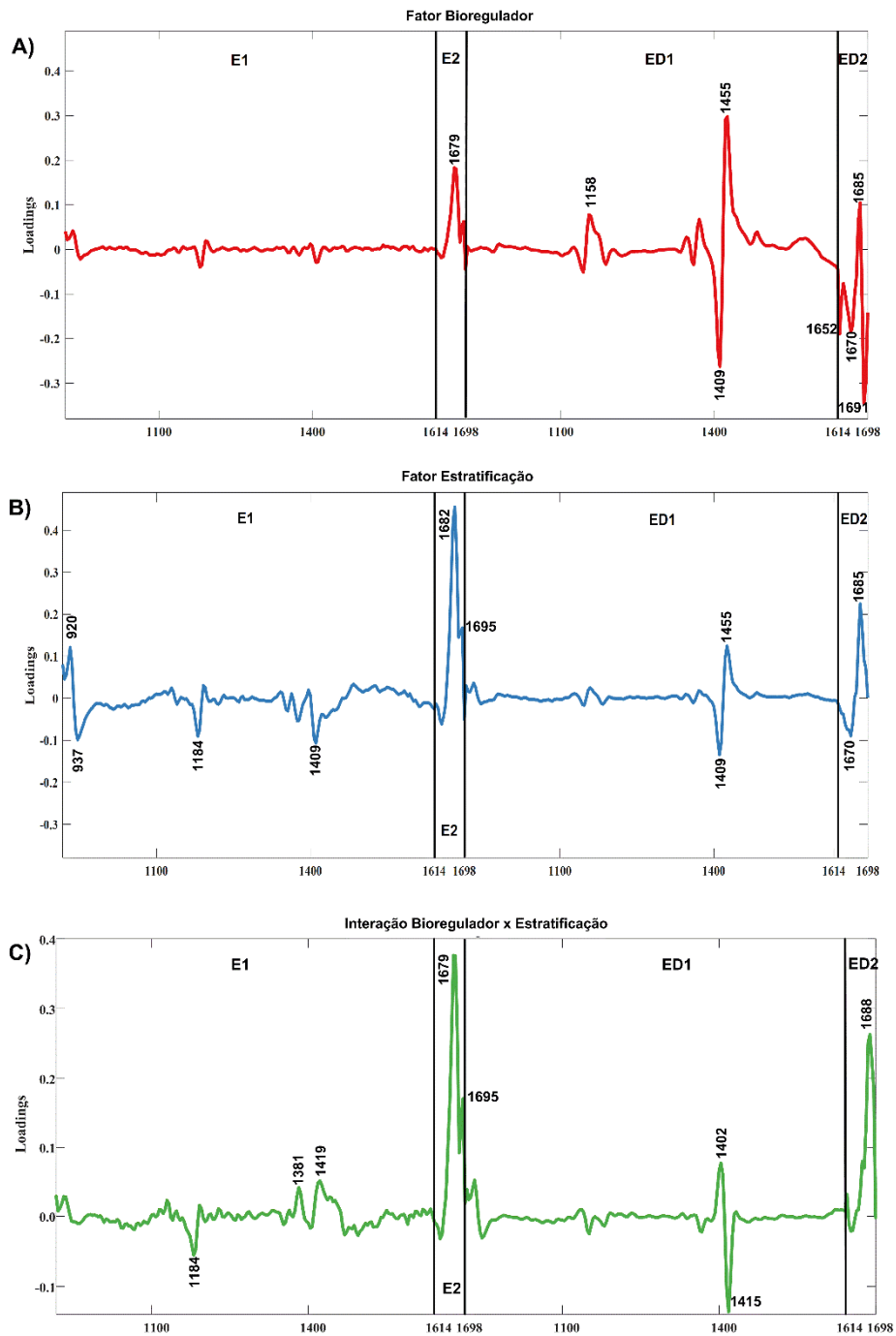


Figura 6. Gráficos dos *loadings* AComDim-ICA das impressões digitais espectrais NIR de grãos de café originados de plantas tratadas com biorreguladores e colhidas nos três estratos verticais: A) Gráfico dos *loadings* CC1×CC3 para fator biorregulador, B) Gráfico dos *loadings* CC1×CC4 para fator de estratificação, C) Gráfico dos *loadings* CC1×CC5 para a interação “biorregulador×estratificação”.

3.2.1 ANÁLISE MULTIBLOCO UV-Vis

As saliências dos 2 blocos para os extratos **E** e **ED** também foram analisadas por AComDim-ICA na região UV-Vis (Fig. 7). CC1, CC4 e CC7 apresentaram valores dos resíduos maior para o conjunto de dados de extratos obtidos pelo solvente **E** (barras vermelhas), e CC2 e CC3 foram os componentes com maior valor de resíduos para **ED** (barras azuis). Os valores F para os blocos de impressão digital espectrais UV foram calculados para cada extrato (**E** e **ED**). Os valores F do conjunto de dados espectrais UV-Vis foram mostrados (Fig. 8), em que o fator de estratificação (STR: bloco 2) teve um valor igual a 1,34 para extratos **E**, enquanto a interação (BIO×STR: bloco) teve um valor de 1,29, ambos superiores ao valor F crítico de 1,21. O conjunto de dados dos extratos **E** apresentou as maiores saliências em CC6 para o fator de estratificação (Fig. 7) e em CC5 para a interação entre os fatores (Fig. 7). Embora a saliência para o fator biorregulador tenha sido maior para os extratos **ED** em CC8 (Fig. 7), não resultou em um valor F significativo (Fig. 8).

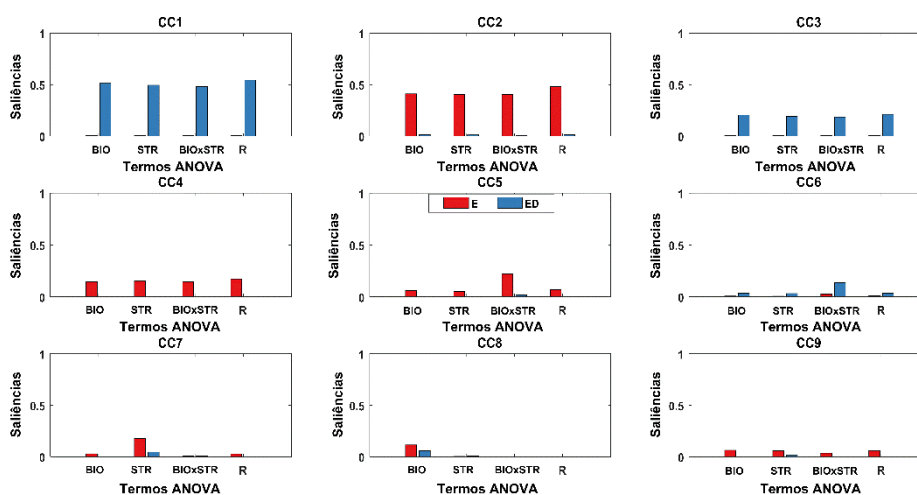


Figura 7. Saliências dos blocos decompostos ANOVA, para Componentes Comuns (CC) 1 a 9 (ver Tabela 1) sendo BIO o fator biorregulador, STR o fator de estratificação, BIO×STR a interação entre biorreguladores e estratificação e, finalmente, R a matriz residual. As barras indicam **E** (etanol - vermelho) e **ED** (etanol e diclorometano - azul).

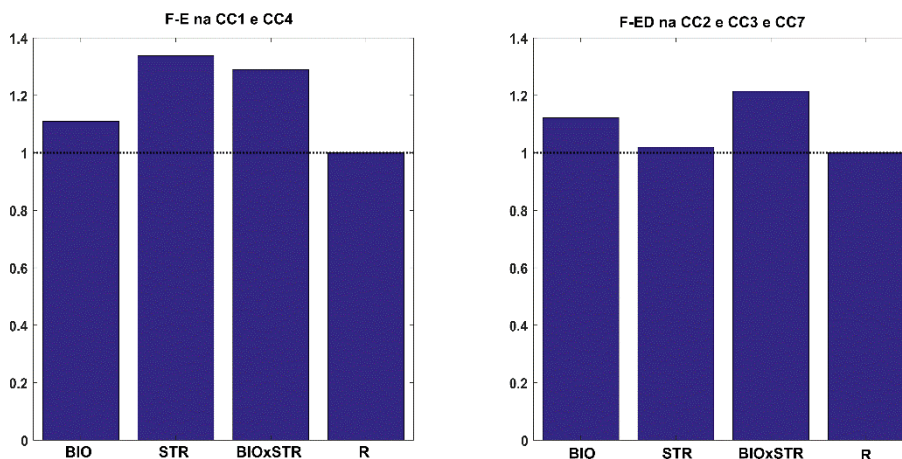


Figura 8. Valores F baseados nos Componentes Comuns (CC) com altas saliências para resíduos calculados a partir do AComDim-ICA de impressões digitais metabolômicas espectrais UV-Vis para grãos de café de plantas tratadas com biorreguladores e colhidas dos três estratos verticais. Sendo, BIO o fator biorregulador, STR o fator de estratificação, BIO×STR a interação entre biorreguladores e estratificação e R a matriz residual.

Os CC5 e CC6 estão relacionados ao fator STR e à interação BIO×STR (Fig. 7). Isso dá origem aos agrupamentos das amostras com base nos níveis do fator STR e na interação BIO×STR apresentado nos gráficos dos escores correspondentes (Fig. 9). Os agrupamentos na CC6×CC1 relacionados aos extratos etanólicos de grãos de café coletados dos três estratos do perfil vertical foram construídos (Fig. 9a). Os grãos do estrato mais baixo da planta apresentaram escores positivos com bandas em 205 e 236 nm de acordo com seus respectivos valores de *loadings* positivos (Fig. 10a). A absorvância espectral na região entre 190 e 210 nm pode ser atribuída a açúcares e lipídios, assim o comprimento de onda em 205 nm pode ser caracterizado por essas substâncias [75], enquanto o comprimento de onda de 236 nm pode ser atribuído a diterpenos. Os grãos dos estratos médio e superior apresentaram valores negativos em CC6, correspondendo a bandas que podem estar relacionadas a compostos como ácido málico e cítrico (214 nm), cafestol (227 nm) e cafeína e/ou trigonelina (260 nm) - [27,76–

78]. Com relação à interação BIO×STR (Fig. 9b), o Ethephon apresentou um impacto divergente no metaboloma do grão em relação ao controle. Isso pode ser observado no CC5, no qual o Ethephon-3 m (Fig. 9b-3) teve valores de escores mais positivos no lado oposto ao controle-3 m (Fig. 9b-6) (experimentos 3 e 6 da Tabela 1, respectivamente), caracterizados por escores negativos (Fig. 9b). Neste caso, os grãos de café Ethephon-3 m (Fig. 9b-3) destacaram bandas em 214, 229, 242 e 260 nm que podem caracterizar a maior concentração relativa de ácido málico e cítrico, cafestol, ácidos clorogênicos e cafeína e/ou trigonelina [22,27,76–78] em relação ao controle-3 m (Fig. 9b-6), conforme mostrado no gráfico dos *loadings* (Fig. 10b). Uma observação semelhante foi feita para o Ethephon-2 m (Fig. 9b-2), caracterizado por *loadings* mais negativos do que o controle-2 m (Fig. 9b-5) (experimentos 2 e 5 da Tabela 1, respectivamente). O gráfico dos *loadings* (Fig. 10b) mostrou bandas (222, 282, 297 e 330 nm), que indicaram maior concentração relativa de ácidos clorogênicos [40,79,80] nos grãos de café Ethephon-2 m (Fig. 9b-2) do que no controle-2 m (Fig. 9b-5).

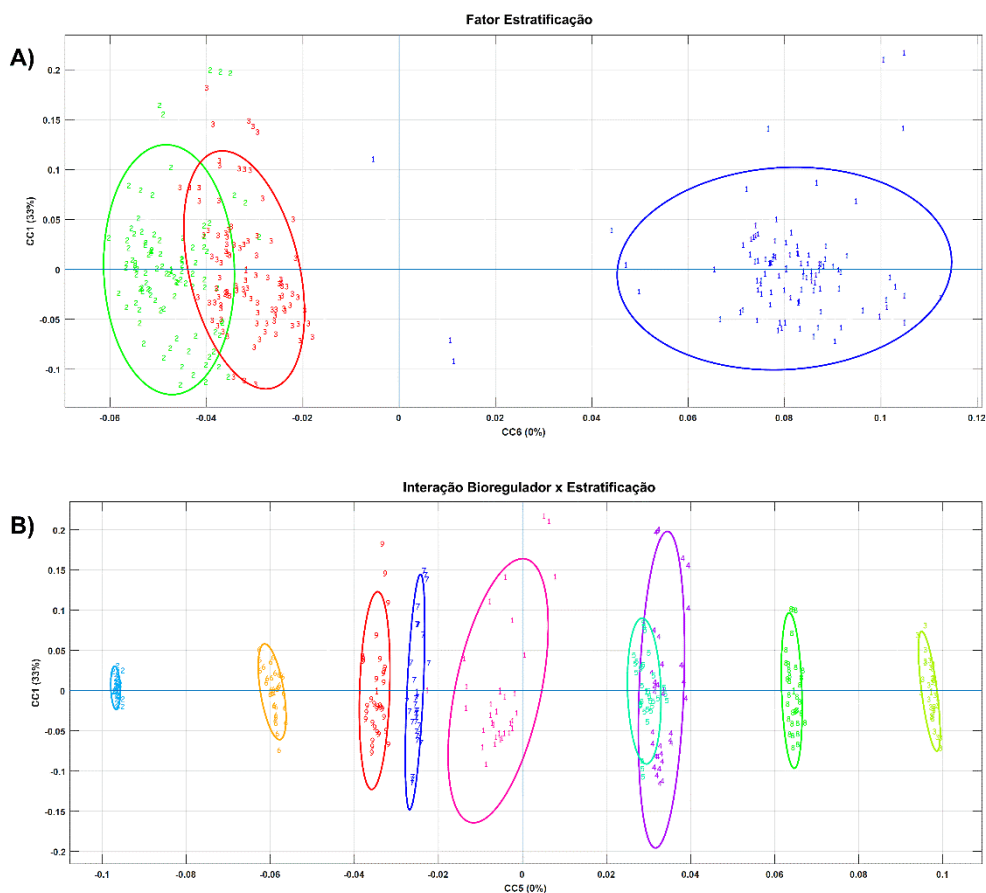


Figura 9. Gráficos dos escores AComDim-ICA das impressões digitais metabolômicas espectrais UV-Vis de grãos de café de plantas tratadas com biorreguladores e colhidas nos três estratos verticais: A) Gráfico dos escores $CC1 \times CC6$ para fator de estratificação. (1) o estrato mais baixo - 1 m, (2) o estrato médio - 2 m e (3) o estrato mais alto - 3 m; B) Gráfico dos escores $CC1 \times CC5$ para interação “biorregulador \times estratificação” - (1) Ethephon-1 m, (2) Ethephon-2 m, (3) Ethephon-3 m, (4) Controle-1 m, (5) Controle-2 m, (6) Controle-3 m, (7) GA3-1 m, (8) GA3-2 m e (9) GA3-3 m.

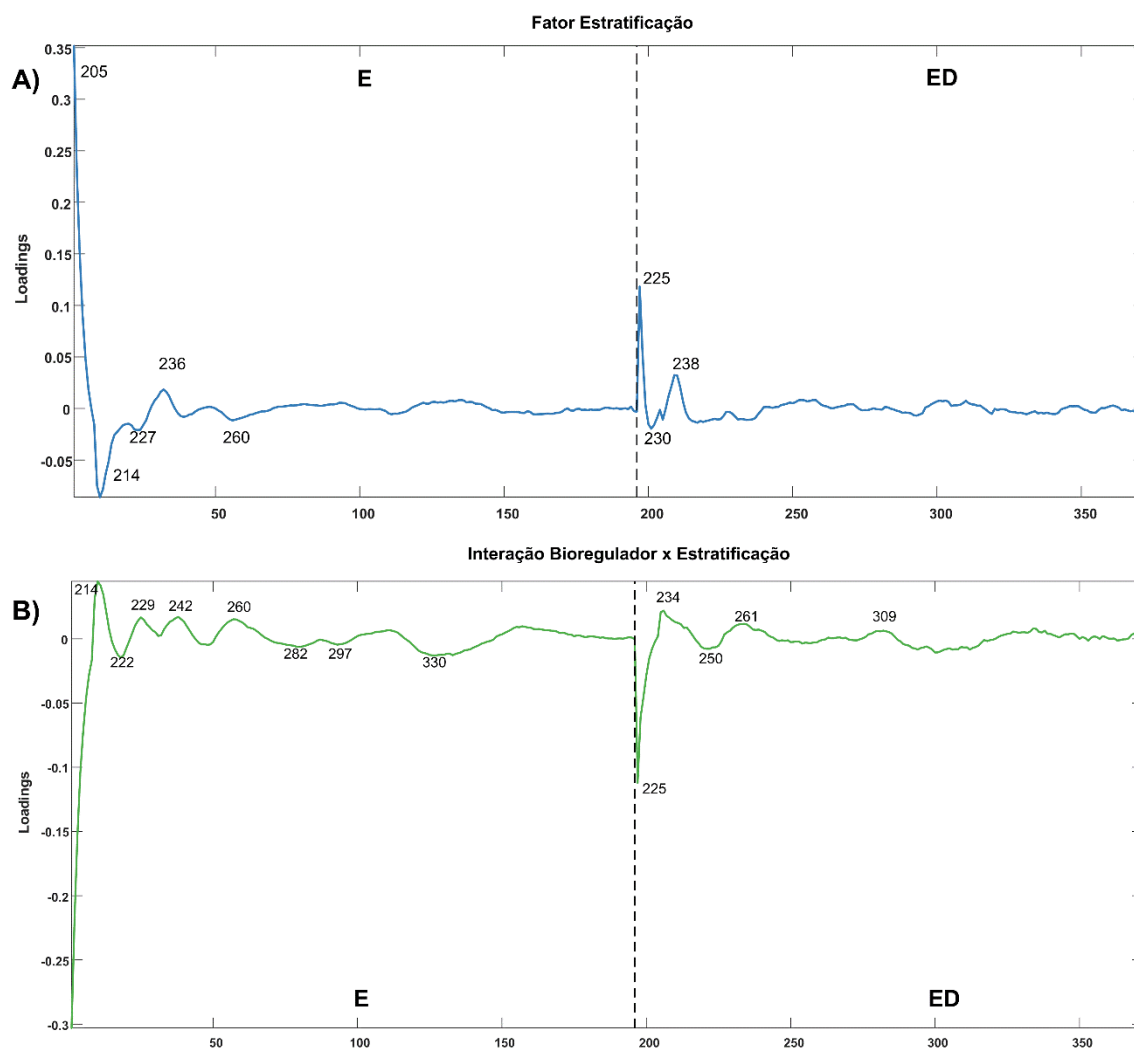


Figura 10. Gráficos dos *loadings* AComDim-ICA das impressões digitais metabolômicas espectrais UV-Vis de grãos de café originados de plantas tratadas com biorreguladores e colhidos nos três estratos verticais de plantas: A) CC1×CC6 gráfico dos *loadings* para fator de estratificação, B) Gráfico dos *loadings* CC1×CC5 para interação “biorregulador×estratificação”.

4 DISCUSSÃO

Como considerações finais, para o fator biorregulador, AComDim-ICA mostrou que as amostras extraídas com o solvente **ED** e analisadas por NIR apresentaram valores de F significativos, enquanto as análises de UV não geraram valores de F significativamente diferentes dos resíduos para os extratos de **ED**. Isso está de acordo com os resultados da Fig. 2A onde a separação das amostras é muito maior para **E** e confirma que a associação do solvente extrator e o método de impressão digital foi fundamental para alcançar resultados satisfatórios. O efeito do GA₃ nas respostas metabólicas foi caracterizado pela maior diferenciação em relação aos efeitos controle ou Ethephon. Conforme indicado pelos *loadings*, os metabólitos que podem estar relacionados a esta resposta foram compostos fenólicos. Alguns estudos mostraram que o uso de GA₃ induz e/ou aumenta a produção de substâncias da classe fenólica em várias espécies vegetais [4]. As substâncias fenólicas estão muito presentes no grão cru, onde são responsáveis pela adstringência do fruto, contribuindo para o sabor e aroma da bebida [81]. Os ácidos clorogênicos foram relacionados aos extratos dos grãos das plantas com aplicação do controle. Quanto maior o teor desses ácidos, menor o grau de amadurecimento do fruto [82,83], o que indica que os grãos coletados de plantas tratadas com GA₃ apresentaram amadurecimento mais avançado em relação aos de plantas de controle ou plantas tratadas com Ethephon.

Os valores F do fator de estratificação foram significativos para as técnicas NIR e UV-Vis quando o solvente **E** foi usado. Os grãos colhidos do estrato mais baixo, de 1 m, foram diferentes dos colhidos nos dois estratos superiores (2 e 3 m), possivelmente apresentando maiores quantidades de cafeína, açúcares, lipídios e diterpenos. Isso pode ser explicado por um efeito de sombreamento nos grãos localizados

no estrato mais baixo, o que pode resultar em maiores teores dessas substâncias do que aqueles que recebem mais luz solar [68,84,85].

Frutos climatéricos são caracterizados por um rápido aumento na biossíntese de etileno, associado a um aumento na taxa de respiração, ocorrendo no início do processo de amadurecimento, que culmina com o amadurecimento dos frutos [86]. O café constitui um fruto climatérico e o etileno desempenha um papel importante no seu processo de amadurecimento, sugerindo que este tipo de fruto é sensível à aplicação de etileno exógeno. A interação entre BIO×STR mostrou que o Ethephon-3 m teve respostas metabolômicas diferentes do controle-3 m. Com a adição de Ethephon, o processo de amadurecimento e alterações químicas podem ter sido afetados, juntamente com uma maior quantidade de cafeína que é observada para estratos vegetais sob luz solar mais intensa [68].

O uso de biorreguladores para auxiliar na sincronização do amadurecimento dos grãos de café acarretou diferenças metabólicas entre os dois hormônios aplicados externamente no experimento, além de serem impactados pela estratificação da planta. As interações entre biorreguladores e condições microambientais dos estratos vegetais demonstraram que a aplicação dos biorreguladores impactou na constituição metabólica do grão, que foi modulada pelas condições microambientais, como luz e temperatura [68].

5 CONCLUSÃO

Alterações metabólicas foram detectadas em grãos crus de *Coffea arabica* tratados com biorreguladores antes de seu amadurecimento final, observados ao longo da estratificação do perfil vertical. A PCA foi capaz de identificar que **E** e **ED** foram os solventes que, entre os aplicados, melhor discriminaram as amostras, tanto para NIR quanto para UV-Vis. O método AComDim-ICA aplicado nos perfis espectrais NIR e UV-Vis dos extratos de grãos de café permitiu avaliar quais os fatores que mais impactaram na composição dos metabólitos. O solvente **ED** apresentou a maior capacidade de extrair metabólitos influenciados pelo efeito biorregulador, enquanto o **E** apresentou a maior capacidade de extrair metabólitos influenciados pelo efeito de estratificação e a interação entre os dois fatores. Além disso, AComDim-ICA mostrou que para o fator BIO compostos fenólicos podem ser responsáveis pela diferenciação entre a aplicação de GA₃ em relação às outras (Ethephon e Controle). Para o fator STR, AComDim-ICA indicou que cafeína, lipídios, açúcares e diterpenos podem ser responsáveis pela diferenciação entre os frutos colhidos no estrato de 1 m em relação aos colhidos nos outros dois estratos (2 m e 3 m). Na interação BIO×STR, a cafeína pode ser a responsável pela diferença entre Ethephon-3 m e controle-3 m. Por fim, o planejamento estatístico de mistura foi essencial para determinar os solventes capazes de extrair os metabólitos que apresentam as diferenças mais significativas em função dos níveis dos fatores. Nosso estudo demonstrou que o uso racional de métodos multivariados e planejamentos experimentais se faz útil para facilitar o estudo do metaboloma de material vegetal.

REFERÊNCIAS

- [1] F.S. Bacilieri, L.C. de Lima, R.M.Q. Lana, D.S.N. Guimarães, M.A. Clemente, Efficacy of a plant growth regulator in the culture of coffee, *Bioscience Journal*. 32 (2016) 346–353. <https://doi.org/10.14393/BJ-v32n2a2016-29820>.
- [2] T.S. Bisht, L. Rawat, B. Chakraborty, V. Yadav, A Recent Advances in Use of Plant Growth Regulators (PGRs) in Fruit Crops - A Review, *International Journal of Current Microbiology and Applied Sciences*. 7 (2018) 1307–1336. <https://doi.org/10.20546/ijemas.2018.705.159>.
- [3] L.M. Moterle, R.F. dos Santos, C.A. Scapim, A. de L. e Braccini, C.M. Bonato, T. Conrado, Efeito de biorregulador na germinação e no vigor de sementes de soja, *Revista Ceres*. 58 (2011) 651–660. <https://doi.org/10.4067/s0718-34292011000300006>.
- [4] K. Jamwal, S. Bhattacharya, S. Puri, Plant growth regulator mediated consequences of secondary metabolites in medicinal plants, *Journal of Applied Research on Medicinal and Aromatic Plants*. 9 (2018) 26–38. <https://doi.org/10.1016/j.jarmap.2017.12.003>.
- [5] R. Wankhade, K. Kubde, MR Deshmukh, G. VV, Effect of bioregulators on yield and yield attributes of Chickpea (*Cicer arietinum* L.), *International Journal of Chemical Studies*. 8 (2020) 140–144. <https://doi.org/10.22271/chemi.2020.v8.i4b.9682>.
- [6] M.R. Ávila, A.D.L. E Braccini, C.A. Scapim, L.P. Albrecht, T.A. Tonin, M. Stülp, Bioregulator application, agronomic efficiency, and quality of soybean seeds, *Scientia Agricola*. 65 (2008) 604–612. <https://doi.org/10.1590/S0103-90162008000600006>.

- [7] N.R. Costa, M. Christian, S. Domingues, J. Domingos, M. Carvalho, M. Teixeira, Performance of coffee plant Icatu red under action of bioregulator applied in the reproductive phases of the culture, *Agrarian*. 2 (2010) 113–130.
- [8] K. Dubravec, I. Dubravec, J. Maniasevic, The effect of the bioregulators agrispon® and ergostim on the vegetative and reproductive growth of apples, *Journal of Sustainable Agriculture*. 5 (1995) 73–83. https://doi.org/10.1300/J064v05n01_06.
- [9] C. Duyvelshoff, J.A. Cline, Ethephon and prohexadione-calcium influence the flowering, early yield, and vegetative growth of young “Northern Spy” apple trees, *Scientia Horticulturae*. 151 (2013) 128–134. <https://doi.org/10.1016/j.scienta.2012.12.002>.
- [10] B. Milić, J. Tarlanović, Z. Keserović, N. Magazin, M. Miodragović, G. Popara, Bioregulators can improve fruit size, yield and plant growth of northern highbush blueberry (*Vaccinium corymbosum* L.), *Scientia Horticulturae*. 235 (2018) 214–220. <https://doi.org/10.1016/j.scienta.2018.03.004>.
- [11] F.B. Silva, C.D. Tormena, E.D. Pauli, A.G. de Almeida, A.B.C. Berg, M. Rakocevic, R.E. Bruns, I.S. Scarminio, G.G. Marcheafave, Time dependent berry maturation for planting density levels in *Coffea arabica* L. beans: Mixture design-fingerprinting using near-infrared transmittance spectroscopy, *Journal of Food Composition and Analysis*. 97 (2021). <https://doi.org/10.1016/j.jfca.2020.103795>.
- [12] F.M. DaMatta, C.P. Ronchi, M. Maestri, R.S. Barros, Ecophysiology of coffee growth and production, *Brazilian Journal of Plant Physiology*. 19 (2007) 485–510.
- [13] G.R. Carvalho, A.N.G. Mendes, L.F. Carvalho, G.F. Bartholo, Eficiência do Ethephon na uniformização e antecipação da maturação de frutos de cafeeiro

- (*Coffea arabica* L.) e na qualidade da bebida, *Ciência e Agrotecnologia*. 27 (2011) 98–106. <https://doi.org/10.1590/s1413-70542003000100012>.
- [14] A.A. Lima, Ethylene regulation under different watering conditions and its possible involvement in coffee (*Coffea arabica* L.) flowering, 2015.
- [15] M.T. Masarirambi, V. Chingwara, V.D. Shongwe, The effect of irrigation on synchronization of coffee (*Coffea arabica* L.) flowering and berry ripening at Chipinge, Zimbabwe, *Physics and Chemistry of the Earth*. 34 (2009) 786–789. <https://doi.org/10.1016/j.pce.2009.06.013>.
- [16] A.M. Al-Saif, A.I. Alebidi, R.S. Al-Obeed, S.S. Soliman, Preharvest Ethephon spray on fruit quality and increasing the rate of ripening of date palm fruit (*Phoenix dactylifera* L.) cv. Helali, *Progress in Nutrition*. 19 (2017) 97–103. <https://doi.org/10.23751/pn.v19i1.4741>.
- [17] J. Kesy, E. Wilmowicz, B. Maciejewska, K. Frankowski, P. Glazinska, J. Kopcewicz, Independent effects of jasmonates and ethylene on inhibition of *Pharbitis nil* flowering, *Acta Physiologiae Plantarum*. 33 (2011) 1211–1216. <https://doi.org/10.1007/s11738-010-0649-9>.
- [18] R. Gupta, S.K. Chakrabarty, Gibberellic acid in plant Still a mystery unresolved, *Plant Signaling & Behavior*. 8 (2013) 1–5.
- [19] M.T. Masarirambi, V.D. Shongwe, V. Chingwara, The effect of GA₃ and ethephon on synchronization of coffee (*Coffea Arabica* L.) flowering and berry ripening, *Acta Horticulturae*. 884 (2010) 573–580. <https://doi.org/10.17660/ActaHortic.2010.884.74>.
- [20] A.K. Kosmides, K. Kamisoglu, S.E. Calvano, S.A. Corbett, I.P. Andoulakis, Metabolomic Fingerprinting: Challenges and Opportunities, *Critical Reviews in Biomedical Engineering*. 41 (2013) 205–221.

- [21] G. Marcheafave, C. Tormena, S. Afonso, M. Rakocevic, R. Bruns, I. Scarminio, Integrated Chemometric Approach to Optimize Sample Preparation for Detecting Metabolic Changes Provoked by Abiotic Stress in *Coffea arabica* L. Leaf Fingerprints, *J Braz Chem Soc.* (2019) 1–10. <https://doi.org/10.21577/0103-5053.20190068>.
- [22] G.G. Marcheafave, E.D. Pauli, C.D. Tormena, L.E. Mattos, A.G. de Almeida, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Irrigated and CO₂ level effects on metabolism in *Coffea arabica* beans from mixture design – near infrared fingerprints, *Microchemical Journal.* 152 (2020) 104276. <https://doi.org/10.1016/j.microc.2019.104276>.
- [23] G.G. Marcheafave, C.D. Tormena, E.D. Pauli, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Experimental mixture design solvent effects on pigment extraction and antioxidant activity from *Coffea arabica* L. leaves, *Microchemical Journal.* 146 (2019) 713–721. <https://doi.org/10.1016/j.microc.2019.01.073>.
- [24] C.D. Tormena, G.G. Marcheafave, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Sequential mixture design optimization for divergent metabolite analysis: Enriched carbon dioxide effects on *Coffea arabica* L. leaves and buds, *Talanta.* 191 (2019) 382–389. <https://doi.org/10.1016/j.talanta.2018.09.002>.
- [25] G.G. Marcheafave, C.D. Tormena, A.E. Terrile, C.A.R. Salamanca-Neto, E.R. Sartori, M. Rakocevic, R.E. Bruns, I.S. Scarminio, E.D. Pauli, Ecometabolic mixture design-fingerprints from exploratory multi-block data analysis in *Coffea arabica* beans from climate changes: Elevated carbon dioxide and reduced soil water availability, *Food Chemistry.* 362 (2021). <https://doi.org/10.1016/j.foodchem.2021.129716>.

- [26] P.H. Hatumura, G.S. de Oliveira, G.G. Marcheafave, M. Rakocevic, R.E. Bruns, I.S. Scarminio, A.E. Terrile, Chemometric Analysis of ^1H NMR Fingerprints of *Coffea arabica* Green Bean Extracts Cultivated under Different Planting Densities, *Food Analytical Methods*. 11 (2018) 1906–1914. <https://doi.org/10.1007/s12161-017-1104-y>.
- [27] A.E. Terrile, G.G. Marcheafave, G.S. Oliveira, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Chemometric Analysis of UV Characteristic Profile and Infrared Fingerprint Variations of *Coffea arabica* Green Beans under Different Space Management Treatments, *J Braz Chem Soc*. 27 (2016) 1254–1263. <https://doi.org/10.5935/0103-5053.20160022>.
- [28] S. Afonso, F. Silva, G. Marcheafave, P. Hatumura, R. Bruns, I. Scarminio, Influence of Seasonality and Sunlight Effects on *Rollinia mucosa* Leaves Fingerprint, *J Braz Chem Soc*. 30 (2018) 968–977. <https://doi.org/10.21577/0103-5053.20180242>.
- [29] M.B. Abreu, G.G. Marcheafave, R.E. Bruns, I.S. Scarminio, M.L. Zeraik, Spectroscopic and Chromatographic Fingerprints for Discrimination of Specialty and Traditional Coffees by Integrated Chemometric Methods, *Food Analytical Methods*. 13 (2020) 2204–2212. <https://doi.org/10.1007/s12161-020-01832-1>.
- [30] G.G. Marcheafave, C.D. Tormena, L.E. Mattos, V.R. Liberatti, A.B.S. Ferrari, M. Rakocevic, R.E. Bruns, I.S. Scarminio, E.D. Pauli, The main effects of elevated CO_2 and soil-water deficiency on ^1H NMR-based metabolic fingerprints of *Coffea arabica* beans by factorial and mixture design, *Science of the Total Environment*. 749 (2020) 142350. <https://doi.org/10.1016/j.scitotenv.2020.142350>.

- [31] I. Moreira, I.S. Scarminio, Chemometric discrimination of genetically modified *Coffea arabica* cultivars using spectroscopic and chromatographic fingerprints, *Talanta*. 107 (2013) 416–422. <https://doi.org/10.1016/j.talanta.2013.01.053>.
- [32] L.M.Z. Garcia, T.F. de Oliveira, P.K. Soares, R.E. Bruns, I.S. Scarminio, Chemometrics and Intelligent Laboratory Systems Statistical mixture design — Principal component determination of synergic solvent interactions for natural product extractions, *Chemometrics and Intelligent Laboratory Systems*. 103 (2010) 1–7. <https://doi.org/10.1016/j.chemolab.2010.04.016>.
- [33] P.K. Soares, I.S. Scarminio, Multivariate chromatographic fingerprint preparation and authentication of plant material from the genus *Bauhinia*, *Phytochemical Analysis*. 19 (2008) 78–85. <https://doi.org/10.1002/pca.1020>.
- [34] P.K. Soares, G.G. Marcheafave, A.A. Gomes, I.S. Scarminio, R.E. Bruns, Mixture design PARAFAC HPLC-DAD metabolomic fingerprints of fractionated organic and basic extracts from *Erythrina speciosa* Andrews leaves, *Chromatographia*. 81 (2018). <https://doi.org/10.1007/s10337-018-3554-9>.
- [35] J.D.P. Lima, A. Farah, B. King, T. de Paulis, P.R. Martin, Distribution of major chlorogenic acids and related compounds in Brazilian green and toasted *Ilex paraguariensis* (Maté) leaves, *Journal of Agricultural and Food Chemistry*. 64 (2016) 2361–2370. <https://doi.org/10.1021/acs.jafc.6b00276>.
- [36] C.D. Tormena, E.D. Pauli, G.G. Marcheafave, G.L. Scheel, M. Rakocevic, R.E. Bruns, I.S. Scarminio, FT-IR biomarkers of sexual dimorphism in yerba-mate plants: Seasonal and light accessibility effects, *Microchemical Journal*. 158 (2020) 105329. <https://doi.org/10.1016/j.microc.2020.105329>.
- [37] S. Afonso, F.B. Silva, G.G. Marcheafave, P.H. Hatumura, R.E. Bruns, I.S. Scarminio, Influence of Seasonality and Sunlight Effects on *Rollinia mucosa*

- Leaves Fingerprint, *J Braz Chem Soc.* 30 (2019) 968–977.
<https://doi.org/10.21577/0103-5053.20180242>.
- [38] R.N.M.J. Páscoa, L.M. Magalhães, J.A. Lopes, FT-NIR spectroscopy as a tool for valorization of spent coffee grounds: Application to assessment of antioxidant properties, *Food Research International.* 51 (2013) 579–586.
<https://doi.org/10.1016/j.foodres.2013.01.035>.
- [39] R.N.M.J. Páscoa, M.C. Sarraguça, L.M. Magalhães, J.R. Santos, A.O.S.S. Rangel, J.A. Lopes, Use of Near-Infrared Spectroscopy for Coffee Beans Quality Assessment, in: *Coffee in Health and Disease Prevention*, Elsevier Inc., 2014: pp. 933–942. <https://doi.org/10.1016/B978-0-12-409517-5.00103-0>.
- [40] A. Belay, A.V. Gholap, Characterization and determination of chlorogenic acids (CGA) in coffee beans by UV-Vis spectroscopy, *African Journal of Pure and Applied Chemistry.* 3 (2009) 234–240. <https://doi.org/10.5897/AJPAC>.
- [41] M. Yulia, D. Suhandy, Indonesian palm civet coffee discrimination using UV-visible spectroscopy and several chemometrics methods, *Journal of Physics: Conference Series.* 835 (2017) 1–7. <https://doi.org/10.1088/1742-6596/835/1/012010>.
- [42] D. Suhandy, M. Yulia, Peaberry coffee discrimination using UV-visible spectroscopy combined with SIMCA and PLS-DA combined with SIMCA and PLS-DA, *International Journal of Food Properties.* 20 (2017) 331–339.
<https://doi.org/10.1080/10942912.2017.1296861>.
- [43] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Variables selection methods in near-infrared spectroscopy, *Analytica Chimica Acta.* 667 (2010) 14–32. <https://doi.org/10.1016/j.aca.2010.03.048>.

- [44] F.C.G.B.S. Alves, P. Valderrama, Ultraviolet spectroscopy and supervised pattern recognition methods for authentication of transgenic and non-transgenic soybean oils, *Analytical Methods*. 7 (2015) 9702–9706. <https://doi.org/10.1039/c5ay02440d>.
- [45] F.S. Grasel, M.C.A. Marcelo, M.F. Ferrão, A non-destructive, rapid and inexpensive methodology based on digital images for the classification of natural tannin extracts, *RSC Advances*. 6 (2016) 32358–32364. <https://doi.org/10.1039/c6ra00900j>.
- [46] I. Marquetti, J.V. Link, A.L.G. Lemes, M.B. dos S. Scholz, P. Valderrama, E. Bona, Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee, *Computers and Electronics in Agriculture*. 121 (2016) 313–319. <https://doi.org/10.1016/j.compag.2015.12.018>.
- [47] L.R. Snyder, P.W. Carr, S.C. Rutan, Solvatochromically based solvent-selectivity triangle, *Journal of Chromatography A*. 656 (1993) 537–547. [https://doi.org/10.1016/0021-9673\(93\)80818-S](https://doi.org/10.1016/0021-9673(93)80818-S).
- [48] N. Christophersen, R.P. Hooper, Multivariate Analysis of Stream Water Chemical Data: The Use of Principal Components Analysis for the End-Member Mixing Problem, *Water Resources Research*. 28 (1992) 99–107.
- [49] G.L. Scheel, E.D. Pauli, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Environmental stress evaluation of *Coffea arabica* L. leaves from spectrophotometric fingerprints by PCA and OSC–PLS–DA, *Arabian Journal of Chemistry*. 12 (2019) 4251–4257. <https://doi.org/10.1016/j.arabjc.2016.05.014>.

- [50] D.J. Bartholomew, Principal Components Analysis, in: International Encyclopedia of Education, 2nd ed., Elsevier, New York, 2010: pp. 374–377. <https://doi.org/10.1016/B978-0-08-044894-7.01358-0>.
- [51] C.D. Tormena, R.C.S. Campos, G.G. Marcheafave, R. Edward Bruns, I.S. Scarminio, E.D. Pauli, Authentication of carioca common bean cultivars (*Phaseolus vulgaris* L.) using digital image processing and chemometric tools, Food Chemistry. 364 (2021) 130349. <https://doi.org/10.1016/j.foodchem.2021.130349>.
- [52] P.D.B. Harrington, N.E. Vieira, J. Espinoza, J.K. Nien, R. Romero, A.L. Yergey, Analysis of variance-principal component analysis: A soft tool for proteomic discovery, Analytica Chimica Acta. 544 (2005) 118–127. <https://doi.org/10.1016/j.aca.2005.02.042>.
- [53] J. Sarembaud, R. Pinto, D.N. Rutledge, M. Feinberg, Application of the ANOVA-PCA method to stability studies of reference materials, Analytica Chimica Acta. 603 (2007) 147–154. <https://doi.org/10.1016/j.aca.2007.09.046>.
- [54] D. Jouan-Rimbaud Bouveresse, R.C. Pinto, L.M. Schmidtke, N. Locquet, D.N. Rutledge, Identification of significant factors by an extension of ANOVA-PCA based on multi-block analysis, Chemometrics and Intelligent Laboratory Systems. 106 (2011) 173–182. <https://doi.org/10.1016/j.chemolab.2010.05.005>.
- [55] M. Meyners, J. Kunert, E.M. Qannari, Comparing generalized procrustes analysis and statis, Food Quality and Preference. 11 (2000) 77–83. [https://doi.org/10.1016/s0950-3293\(99\)00038-5](https://doi.org/10.1016/s0950-3293(99)00038-5).
- [56] R. Karoui, É. Dufour, J. de Baerdemaeker, Common components and specific weights analysis: A tool for monitoring the molecular structure of semi-hard

- cheese throughout ripening, *Analytica Chimica Acta*. 572 (2006) 125–133.
<https://doi.org/10.1016/j.aca.2006.04.089>.
- [57] V. Cariou, E.M. Qannari, D.N. Rutledge, E. Vigneau, ComDim: From multiblock data analysis to path modeling, *Food Quality and Preference*. 67 (2018) 27–34.
<https://doi.org/10.1016/j.foodqual.2017.02.012>.
- [58] J.N. Miller, J.C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, 6th ed., Pearson Education Limited, Harlow, UK, 2010.
<https://doi.org/10.7861/clinmedicine.14-6-677>.
- [59] N. Delcourt, C. Rébufa, N. Dupuy, N. Boukhdoud, C. Brunel, J. Abadie, I. Giffard, A.M. Farnet-Da Silva, Infrared spectroscopy as a useful tool to predict land use depending on Mediterranean contrasted climate conditions: A case study on soils from olive-orchards and forests, *Science of the Total Environment*. 686 (2019) 179–190. <https://doi.org/10.1016/j.scitotenv.2019.05.240>.
- [60] M. de Figueiredo, S. Giannoukos, C. Wüthrich, R. Zenobi, D.N. Rutledge, A tutorial on the analysis of multifactorial designs from one or more data sources using AComDim, *Journal of Chemometrics*. Accepted (2021) 1–27.
- [61] M. Hanafi, E.M. Qannari, Nouvelles Propriétés De L'Analyse En Composantes Communes Et Poids Spécifiques, *Journal de La Société Française de Statistique*. 149 (2008) 75–97.
- [62] D.N. Rutledge, D. Jouan-Rimbaud Bouveresse, Independent Components Analysis with the JADE algorithm, *TrAC - Trends in Analytical Chemistry*. 50 (2013) 22–32. <https://doi.org/10.1016/j.trac.2013.03.013>.
- [63] D.N. Rutledge, D. Jouan-Rimbaud Bouveresse, Corrigendum to “Independent Components Analysis with the JADE algorithm,” *TrAC - Trends in Analytical Chemistry*. 67 (2015) 220. <https://doi.org/10.1016/j.trac.2015.02.001>.

- [64] D.N. Rutledge, Comparison of Principal Components Analysis, Independent Components Analysis and Common Components Analysis, *Journal of Analysis and Testing*. 2 (2018) 235–248. <https://doi.org/10.1007/s41664-018-0065-5>.
- [65] C.B.Y. Cordella, D. Bertrand, SAISIR: A new general chemometric toolbox, *TrAC - Trends in Analytical Chemistry*. 54 (2014) 75–82. <https://doi.org/10.1016/j.trac.2013.10.009>.
- [66] J.F. Cardoso, A. Souloumiac, Blind beamforming for non-Gaussian signals, *Proceedings of the Institution of Electrical Engineers*. 140 (1993) 62–70.
- [67] M. Rakocevic, E.R. Batista, R.A.A. Pazianotto, M.B.S. Scholz, G.A.R. Souza, E. Campostrini, J.C. Ramalho, Leaf gas exchange and bean quality fluctuations over the whole canopy vertical profile of Arabica coffee cultivated under elevated CO₂, *Functional Plant Biology*. 48 (2021) 469–482.
- [68] M. Rakocevic, F.T. Matsunaga, D.F. Baroni, E. Campostrini, E. Costes, Multiscale analyses of growth and berry distributions along four branching orders and vertical profile of *Coffea arabica* L. cultivated under high-density planting systems, *Scientia Horticulturae*. 281 (2021) 109934. <https://doi.org/10.1016/j.scienta.2021.109934>.
- [69] L.E. Rodriguez-Saona, F.S. Fry, M.A. McLaughlin, E.M. Calvey, Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy, *Carbohydrate Research*. 336 (2001) 63–74. [https://doi.org/10.1016/S0008-6215\(01\)00244-0](https://doi.org/10.1016/S0008-6215(01)00244-0).
- [70] J.Y. Chen, H. Zhang, R. Matsunaga, Visible and Near Infrared Spectroscopy for Rapid Analysis of the Sugar Composition of Raw Ume Juice, *Food Science and Technology Research*. 13 (2007) 291–295. <https://doi.org/10.3136/fstr.13.291>.
- [71] A. Bedini, V. Zanolli, S. Zanardi, U. Bersellini, E. Dalcanale, M. Suman, Rapid and Simultaneous Analysis of Xanthines and Polyphenols as Bitter Taste Markers

- in Bakery Products by FT-NIR Spectroscopy, *Food Analytical Methods*. 6 (2013) 17–27. <https://doi.org/10.1007/s12161-012-9405-7>.
- [72] C. Pizarro, I. Esteban-Díez, J.-M. González-Sáiz, M. Forina, Use of Near-Infrared Spectroscopy and Feature Selection Techniques for Predicting the Caffeine Content and Roasting Color in Roasted Coffees, *Journal of Agricultural and Food Chemistry*. 55 (2007) 7477–7488. <https://doi.org/10.1021/jf071139x>.
- [73] S. Buratti, N. Sinelli, E. Bertone, A. Venturello, E. Casiraghi, F. Geobaldo, Discrimination between washed Arabica, natural Arabica and Robusta coffees by using near infrared spectroscopy, electronic nose and electronic tongue analysis, *Journal of the Science of Food and Agriculture*. 95 (2015) 2192–2200. <https://doi.org/10.1002/jsfa.6933>.
- [74] L.M. Magalhães, S. Machado, M.A. Segundo, J.A. Lopes, R.N.M.J. Páscoa, Rapid assessment of bioactive phenolics and methylxanthines in spent coffee grounds by FT-NIR spectroscopy, *Talanta*. 147 (2016) 460–467. <https://doi.org/10.1016/j.talanta.2015.10.022>.
- [75] F. Delarozza, M. Rakocevic, G.B. Malta, R.E. Bruns, I.S. Scarminio, Spectroscopic and Chromatographic Fingerprint Analysis of Composition Variations in *Coffea arabica* Leaves Subject to Different Light Conditions and Plant Phenophases, *J Braz Chem Soc*. 25 (2014) 1929–1938.
- [76] M. Yulia, D. Suhandy, Identification of fresh and expired ground roasted robusta coffee using UV-visible spectroscopy and chemometrics, *MATEC Web of Conferences*. 197 (2018) 4–8. <https://doi.org/10.1051/matecconf/201819709003>.
- [77] D.L. Kalschne, M.C. Viegas, A.J. de Conti, M.P. Corso, M. de T. Benassi, Steam pressure treatment of defective *Coffea canephora* beans improves the volatile profile and sensory acceptance of roasted coffee blends, *Food Research*

- International. 105 (2018) 393–402.
<https://doi.org/10.1016/j.foodres.2017.11.017>.
- [78] D. Habtamu, A. Belay, First order derivative spectra to determine caffeine and chlorogenic acids in defective and nondefective coffee beans, *Food Science and Nutrition*. 8 (2020) 4757–4762. <https://doi.org/10.1002/fsn3.1723>.
- [79] G. Navarra, M. Moschetti, V. Guarrasi, M.R. Mangione, V. Militello, M. Leone, Simultaneous determination of caffeine and chlorogenic acids in green coffee by UV/Vis spectroscopy, *Journal of Chemistry*. 2017 (2017) 1–8.
<https://doi.org/10.1155/2017/6435086>.
- [80] J.S. da Silveira, C. Mertz, G. Morel, S. Lacour, M.P. Belleville, N. Durand, M. Dornier, Alcoholic fermentation as a potential tool for coffee pulp detoxification and reuse: Analysis of phenolic composition and caffeine content by HPLC-DAD-MS/MS, *Food Chemistry*. 319 (2020).
<https://doi.org/10.1016/j.foodchem.2020.126600>.
- [81] F.A. de Sousa, Avaliação de polifenóis e atividade antioxidante do café arábica (*Coffea arabica*), 2017.
- [82] M. Rakocevic, M.B. dos S. Scholz, C.S.G. Kitzberger, Berry distributions on coffee trees cultivated under high densities modulate the chemical composition of respective coffee beans during one biannual cycle, *International Journal of Fruit Science*. 18 (2018) 117–137. <https://doi.org/10.1080/15538362.2017.1422448>.
- [83] G.G. Marcheafave, E.D. Pauli, C.D. Tormena, M.C. v. Ortiz, A.G. de Almeida, M. Rakocevic, R.E. Bruns, I.S. Scarminio, Factorial design fingerprint discrimination of *Coffea arabica* beans under elevated carbon dioxide and limited water conditions, *Talanta*. 209 (2020) 120591.
<https://doi.org/10.1016/j.talanta.2019.120591>.

- [84] K. Tolessa, J. D'heer, L. Duchateau, P. Boeckx, Influence of growing altitude, shade and harvest period on quality and biochemical composition of Ethiopian specialty coffee, *Journal of the Science of Food and Agriculture*. 97 (2017) 2849–2857. <https://doi.org/10.1002/jsfa.8114>.
- [85] L. Tuccio, P. Pinelli, A. Godina, J.M. Medina, G. Agati, The chemical composition and quality of the Parainema coffee cultivar under different shading conditions, as assessed by a leaf flavonol optical index, *European Food Research and Technology*. 245 (2019) 1733–1741. <https://doi.org/10.1007/s00217-019-03285-1>.
- [86] S.A. Sagio, A.A. Lima, H.G. Barreto, C.H.S. de Carvalho, L.V. Paiva, A. Chalfun-Junior, Physiological and molecular analyses of early and late *Coffea arabica* cultivars at different stages of fruit ripening, *Acta Physiologiae Plantarum*. 35 (2013) 3091–3098. <https://doi.org/10.1007/s11738-013-1342-6>.

CONCLUSÃO GERAL

A aplicação de métodos quimiométricos em impressões digitais de erva-mate obtidas através da espectroscopia de infravermelho pode identificar possíveis biomarcadores sexuais independente da fenofase e de níveis de luz. Os biomarcadores FT-IR de dimorfismo sexual permaneceram constantes independente da fenofase (primavera e inverno) e variações de luminosidade (sombra e clareira). Dessa forma, folhas de erva-mate apresentaram a cafeína como um possível metabólito biomarcador para plantas masculinas e a feofitina *a* para as fêmeas.

O uso de imagens digitais com seus histogramas de cor para autenticação de feijões só foi possível através da utilização do método quimiométrico PLS-DA. Os modelos construídos para os três cultivares indicaram bom ajuste de classificação e predição, podendo ser explorado na análise de controle de qualidade desses grãos.

Através do método AComDim-ICA foi possível detectar nos perfis espectrais NIR e UV-Vis alterações metabólicas nos grãos crus de *Coffea arabica* tratados com biorreguladores ao longo do perfil vertical da planta. Para o fator biorregulador compostos fenólicos podem ser responsáveis pela diferença entre plantas com aplicação de GA₃ em relação as outras (Ethephon e Controle). Para o fator estratificação cafeína, lipídios, açúcares e diterpenos podem ser responsáveis pela diferença dos frutos colhidos a 1 metro do chão em relação aos colhidos nos outros dois estratos.

Em suma, o trabalho mostra que o uso racional de métodos quimiométricos se faz útil para estudar diferentes produtos agrícolas sem a necessidade de instrumentação analítica sofisticada e onerosas. Além disso, ele torna fácil a interpretação de dados volumosos obtidos durante as análises.

ANEXOS

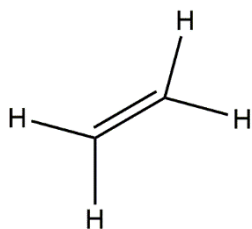


Figura A4: Estrutura molecular etileno, produzido nas plantas naturalmente ou após adição de biorregulador.

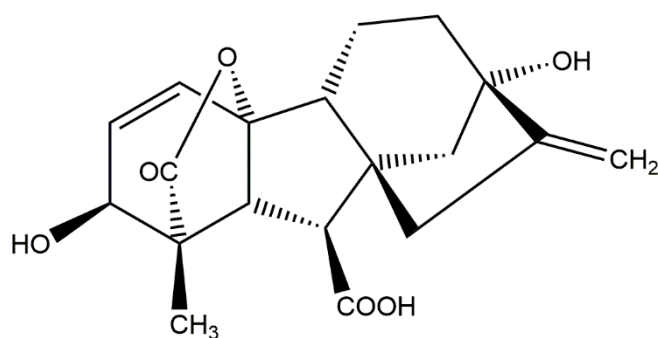


Figura A5: Estrutura molecular do ácido giberélico GA₃.

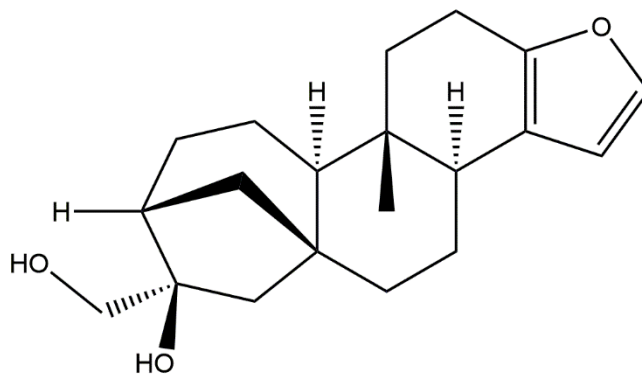


Figura A6: Estrutura molecular do cafestol.

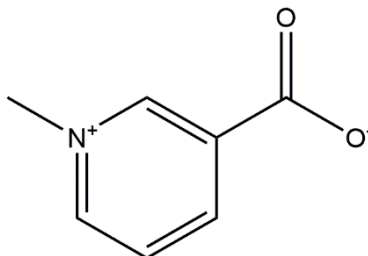
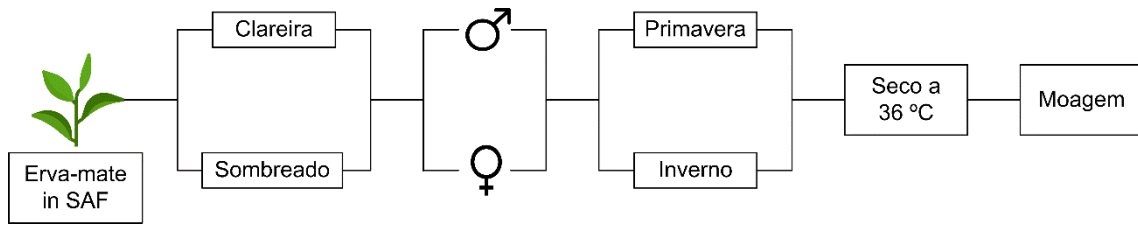
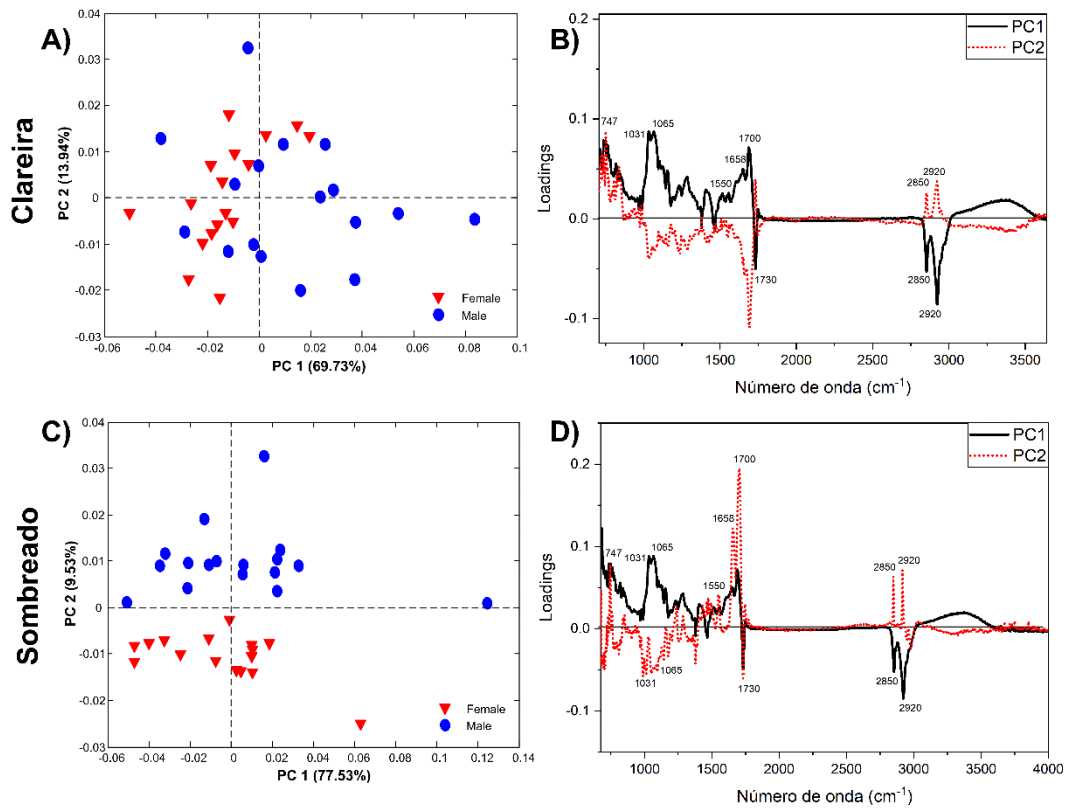


Figura A7: Estrutura molecular da trigonelina.

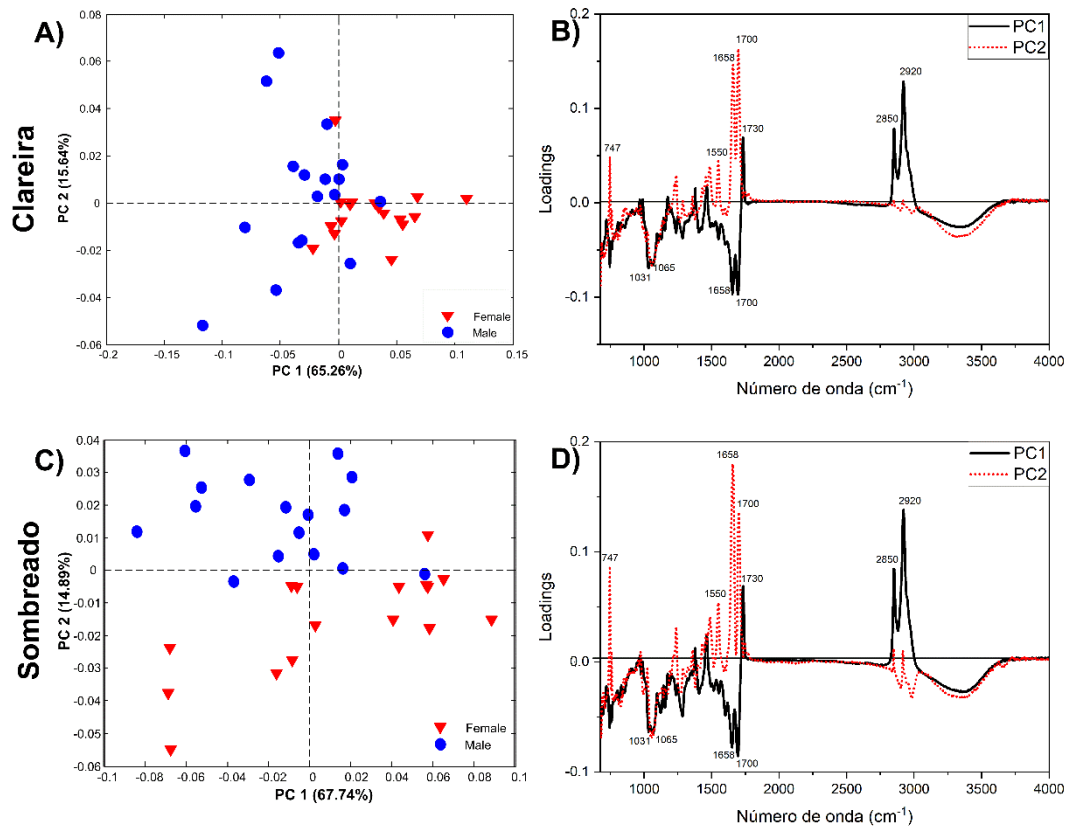
ANEXO B
Figuras em alta qualidade

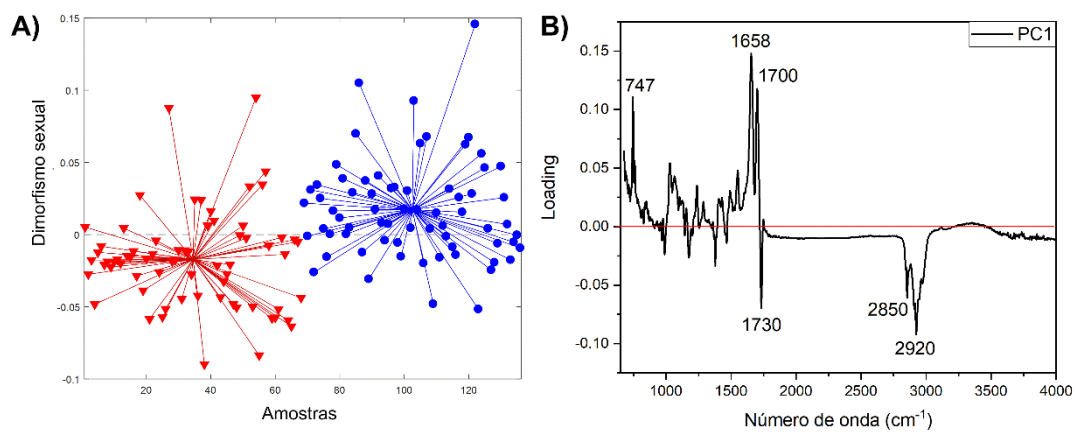


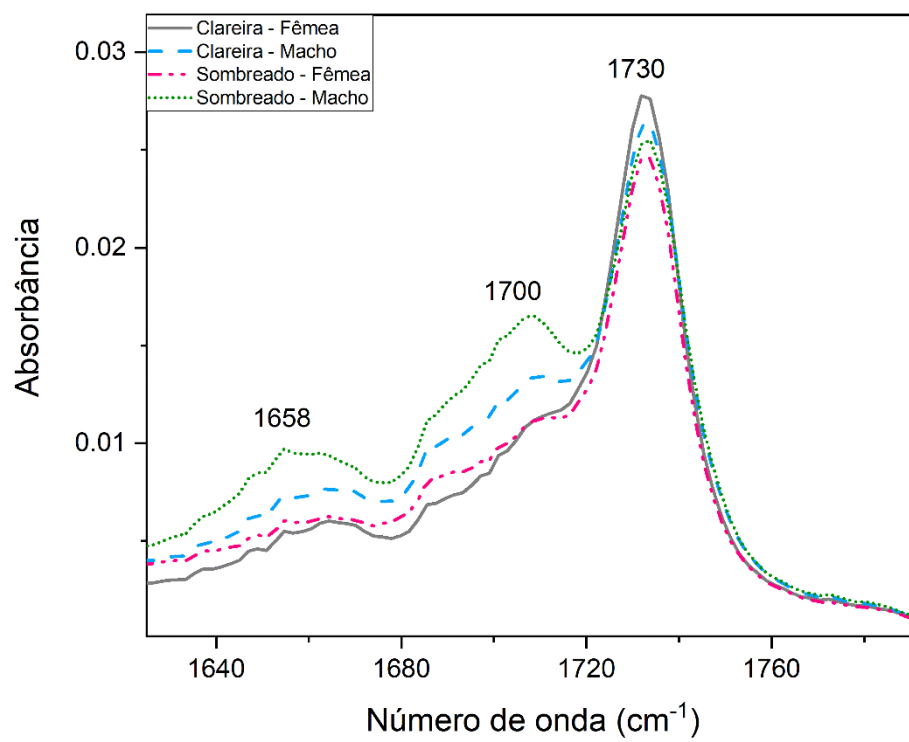
Primavera

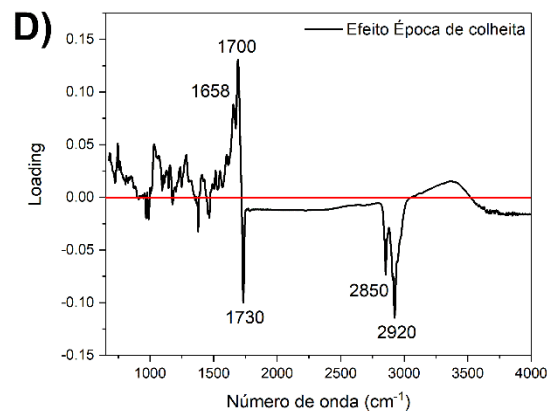
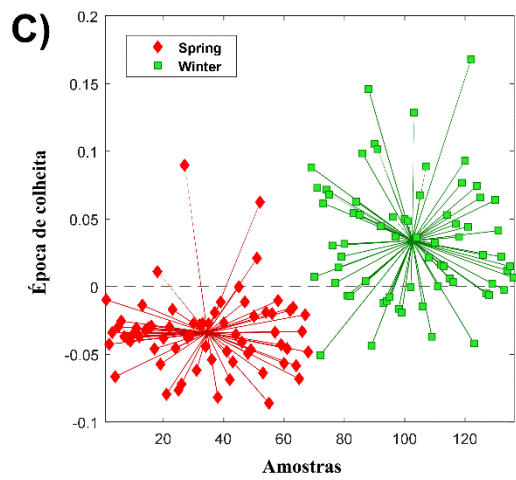
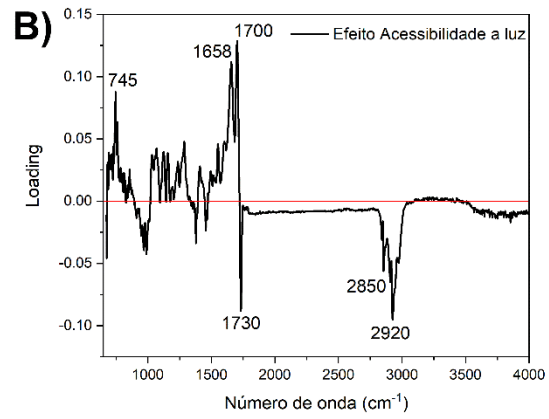
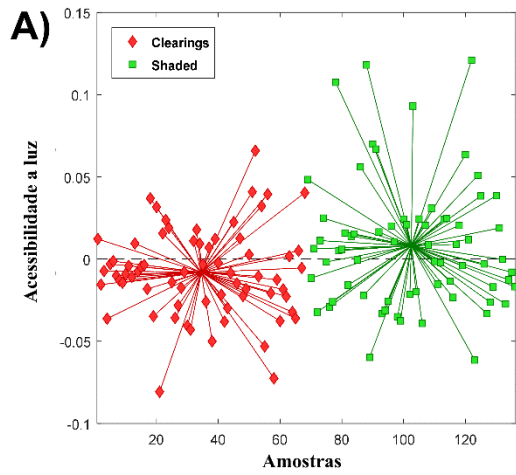


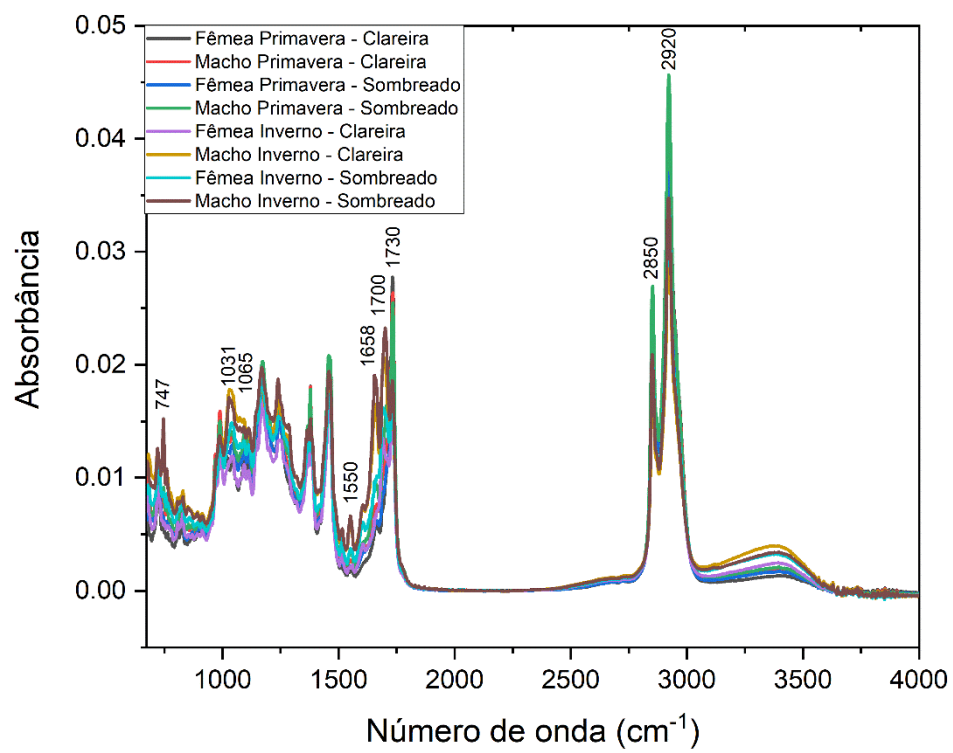
Inverno

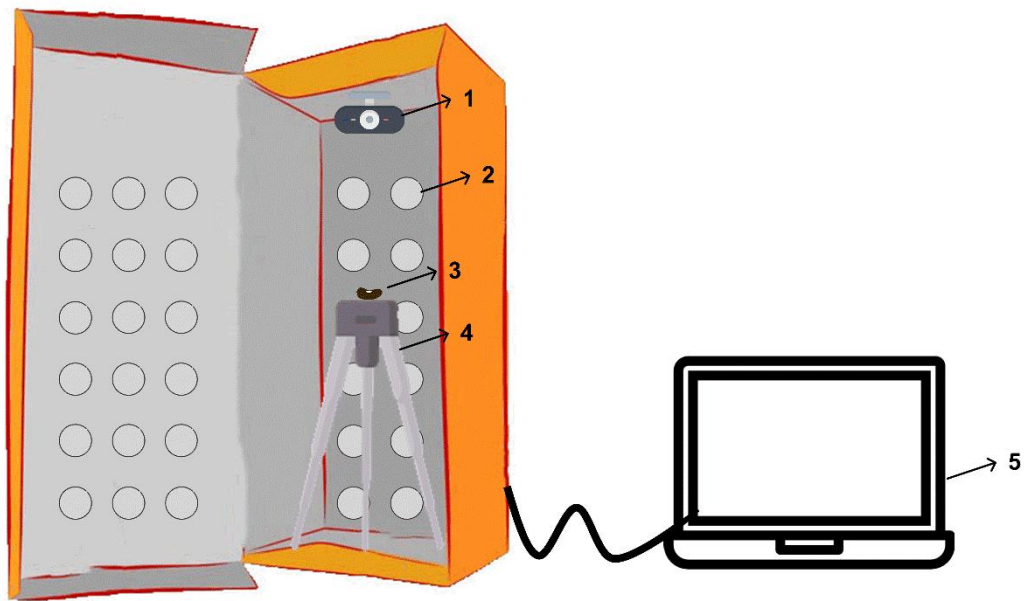


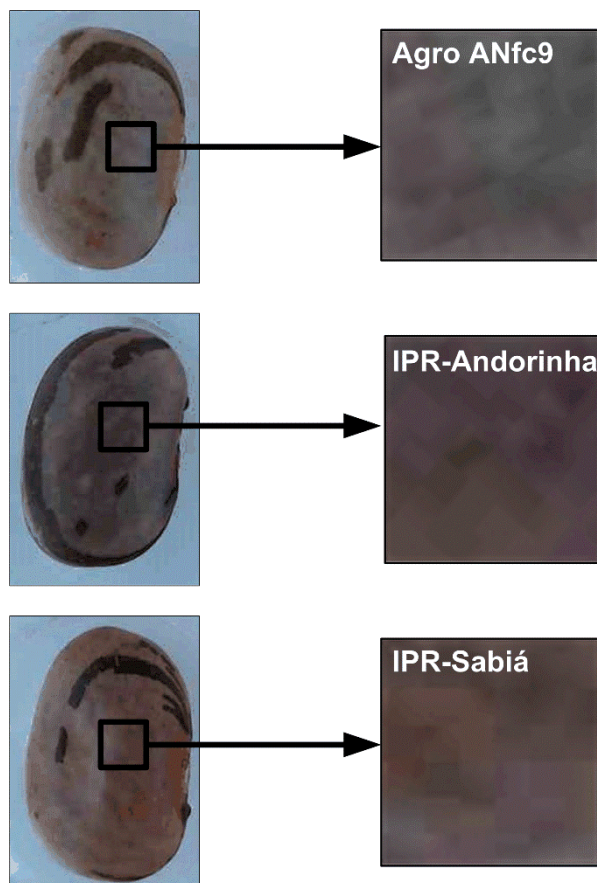




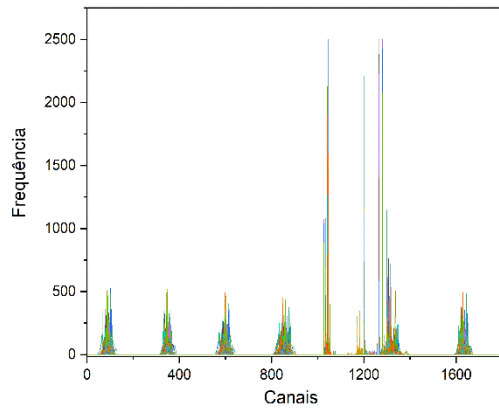




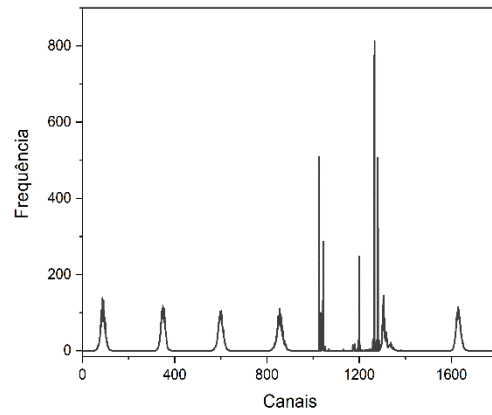




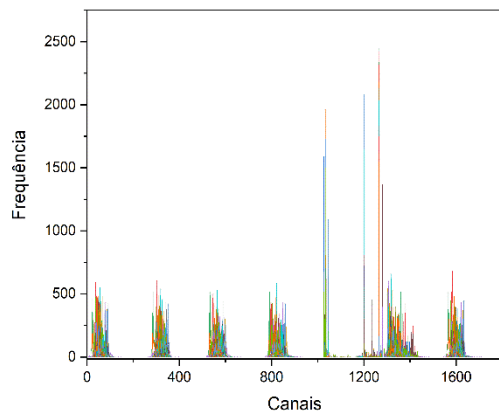
A) Agro ANfc9



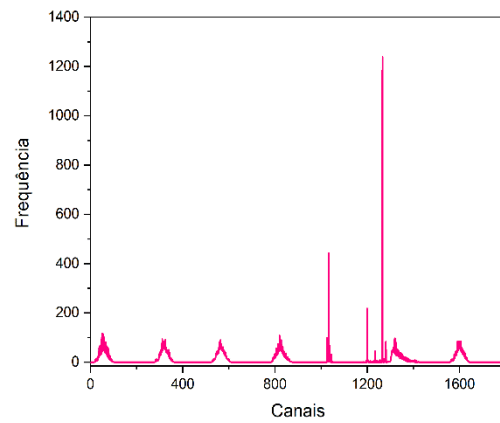
B) Perfil médio Agro ANfc9



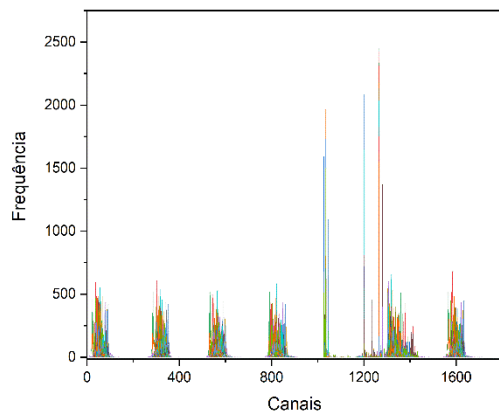
C) IPR-Andorinha



D) Perfil médio IPR-Andorinha



E) IPR-Sabiá



F) Perfil médio IPR-Sabiá

