



UNIVERSIDADE
ESTADUAL DE LONDRINA

ANDRESSA GOMES DE ALMEIDA

**MÉTODOS QUIMIOMÉTRICOS ALIADOS À
ESPECTROSCOPIA NIR PARA O MONITORAMENTO DE
FOLHAS DE ERVA-MATE QUIMICAMENTE
DIFERENCIADAS**

Londrina
2023

ANDRESSA GOMES DE ALMEIDA

**MÉTODOS QUIMIOMÉTRICOS ALIADOS À
ESPECTROSCOPIA NIR PARA O MONITORAMENTO DE
FOLHAS DE ERVA-MATE QUIMICAMENTE
DIFERENCIADAS**

Dissertação apresentada à Universidade Estadual de Londrina - UEL, como requisito parcial para a obtenção do título de Mestre em Química Analítica

Orientadora: Prof^a. Ieda Spacino Scarmínio

Londrina
2023

A447m Almeida, Andressa Gomes de.

Métodos quimiométricos aliados à espectroscopia NIR para o monitoramento de folhas de erva-mate quimicamente diferenciadas / Andressa Gomes de Almeida. - Londrina, 2023.

55 f. : il.

Orientador: Ieda Spacino Scarminio

Dissertação (Mestrado em Química) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Química, 2023.

Inclui bibliografia.

1. Quimiometria - Tese. 2. Espectroscopia no infravermelho próximo - Tese. 3. Erva-mate - Tese. I. Scarminio, Ieda Spacino. II. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Química. III. Título.

CDU 54

ANDRESSA GOMES DE ALMEIDA

**MÉTODOS QUIMIOMÉTRICOS ALIADOS À
ESPECTROSCOPIA NIR PARA O MONITORAMENTO DE
FOLHAS DE ERVA-MATE QUIMICAMENTE
DIFERENCIADAS**

Dissertação apresentada à Universidade Estadual de Londrina - UEL, como requisito parcial para a obtenção do título de Mestre em Química Analítica.

BANCA EXAMINADORA

Prof^ª. Dr^ª. Ieda Spacino Scarmínio
Universidade Estadual de Londrina - UEL

Prof^ª. Dr^ª. Marcela Zanetti Corazza
Universidade Estadual de Londrina - UEL

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

Londrina, 02 de Outubro de 2023.

RESUMO

Almeida, Andressa Gomes de. **Métodos quimométricos aliados à espectroscopia NIR para monitoramento de folhas de erva-mate quimicamente diferenciadas.** 2023. 55. Dissertação de Mestrado (Química) – Centro de Ciências Exatas, Universidade Estadual de Londrina, Londrina, 2023.

Insumos de erva-mate (*Ilex paraguariensis* St. Hil., Aquifoliaceae) têm sido utilizados em diversos setores da indústria. Quando ingerida, promove efeitos benéficos como prevenção de cânceres e doenças cardiovasculares. A cafeína, presente na erva-mate, quando consumida moderadamente traz diversos benefícios, entretanto, o abuso pode levar a danos ao consumidor. É conhecido que fatores ambientais interferem nas rotas biossintéticas das plantas. Pesquisas recentes têm utilizado esta informação em favor da produção de plantas com conteúdos diversificados de metabólitos, como a cafeína. Dois grupos de clones de erva-mate geneticamente melhorados foram classificados como: plantas com alto teor de cafeína e descafeinadas. As plantas de erva-mate foram cultivadas no sistema semi-hidropônico sob cinco níveis de sombreamento (0%, 40%, 51%, 76% e 82%). Impressões digitais obtidas por meio da espectroscopia de infravermelho próximo (NIRS) foram obtidas para as folhas de erva-mate quimicamente diferenciadas e modelos de análise de variância - análise de componentes simultâneos (ASCA) foram desenvolvidos para monitoramento químico, onde o pré-processamento dos espectros NIR foi avaliado através de um planejamento fatorial completo de três níveis. O tratamento por meio da Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA) em dados de refletância (900–1700 nm) obtidos pelo NIR, foi usado para identificar as folhas de erva-mate, com análise direta no material vegetal, bem como a análise de componentes principais (PCA) foi aplicada para avaliar os efeitos da luminosidade sobre as plantas. Ambos tratamentos estatísticos trouxeram resultados satisfatórios, indicando o uso das técnicas como alternativa a processos mais dispendiosos, de maior custo relativo e geradores de resíduos tóxicos.

Palavras-chave: Erva-mate geneticamente melhorada; Cafeína, Condições de luz, NIRS, PLS-DA, PCA, ASCA, Planejamento fatorial.

ABSTRACT

Almeida, Andressa Gomes de. **Chemometric methods allied to NIR spectroscopy for monitoring chemically differentiated yerba mate leaves**. 2023. 55. Dissertação de Mestrado (Química) – Centro de Ciências Exatas, Universidade Estadual de Londrina, Londrina, 2023.

Inputs from yerba mate (*Ilex paraguariensis* St. Hil., Aquifoliaceae) have been used in several industrial sectors. When ingested, it promotes beneficial effects such as preventing cancer and cardiovascular diseases. Caffeine, present in yerba mate, when consumed moderately brings several benefits, however, abuse can lead to serious damage at the consumer. It is known that environmental factors interfere in the biosynthetic routes of plants. Recent research has used this information in favor of the production of plants with diverse metabolite contents, such as caffeine. Two groups of genetically improved yerba mate clones were classified as: plants with high caffeine content and decaffeinated. Yerba mate plants were grown in a semi-hydroponic system under five shading levels (0%, 40%, 51%, 76% and 82%). Fingerprints obtained using Near Infrared Spectroscopy (NIRS) were obtained for chemically differentiated yerba mate leaves and models of variance analysis - ANOVA – Simultaneous Component Analysis, (ASCA) were developed for chemical monitoring, where the pre-processing of NIR spectra was evaluated through a three - level full factorial design. Treatment using Partial Least Squares-Discriminant Analysis (PLS-DA) on reflectance data (900–1700 nm) obtained by NIR was used to identify yerba mate leaves, with direct analysis on plant material, as well as principal component analysis (PCA) was applied to evaluate the effects of light on plants. Both statistical treatments brought satisfactory results, indicating the use of the techniques as an alternative to more expensive processes, with higher relative costs and generating toxic waste.

Keywords: Genetically improved yerba mate; Caffeine, Light conditions, NIRS, PLS-DA, PCA, ASCA, Factorial design.

SUMÁRIO

Resumo	5
Abstract	6
Introdução	7
Referências	9
Determinação Espectral NIR Direta de Melhoramento Genético, Disponibilidade Luminosa e seus efeitos de interação na Composição Química de Folhas de Erva-Mate Quimicamente Seleccionadas	11
Resumo	12
1.Introdução	13
2.Materiais e Métodos	16
2.1 Produção da planta e amostragem	16
2.2 Medições NIR	16
2.3 Tratamentos quimiométricos	17
2.3.1 Efeitos de pré-processamento em resultados ASCA.....	17
2.3.2 Modelos ASCA	18
3.Resultados e Discussão	19
Conclusão	27
Agradecimentos	27
Referências	28
Material Suplementar	31
Uso da Espectroscopia NIR portátil para identificação de clones de Erva-mate (Ilex Paraguariensis) quimicamente diferenciados	33
Resumo	34
1.Introdução	35
2.Materiais e Métodos	37
2.1 Produção da planta e amostragem	37
2.2 Medições NIR	38
2.3 Tratamentos Quimiométricos	38
2.3.1 Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA).....	38
2.3.2 Análise de componentes principais (PCA).....	39
2.4 Software e Análises estatísticas	39
3.Resultados e Discussão	39
Conclusão	47
Agradecimentos	48
Referências	49
Material Suplementar	53

Introdução

A erva-mate (*Ilex paraguariensis* St. Hil., Aquifoliaceae) é mundialmente conhecida devido ao seu consumo sob as formas de chimarrão, tererê e chá-mate. Sua composição química rica em bioativos como: alcalóides, polifenóis, saponinas, óleos essenciais, vitaminas e minerais, que possuem atividades estimulantes, diuréticas, associadas à prevenção de cânceres, danos no DNA e doenças cardiovasculares (Mazur et al., 2014) além de promoverem ação antioxidante, antimicrobiana (Oh et al., 2013) e antibacteriana (Fayad et al., 2020), tem atraído a atenção de diversos setores da indústria e pesquisa.

É possível encontrar a erva-mate cultivada sob os sistemas de monocultura e agroflorestal, conseqüentemente variações morfológicas (Rakocevic & Martim, 2011), fisiológicas (Rakocevic et al., 2011; Takeshi Matsunaga et al., 2014) e metabólicas (Tormena et al., 2020) ocorrem devido a variada incidência luminosa. Alguns estudos indicam variações no teor de metabólitos secundários da erva-mate devido às condições ambientais, como clima, variabilidade genética, tipo de cultivo, idade da planta e condições agrônômicas (Pauli et al., 2019; Tormena et al., 2020).

A cafeína é um metabólito secundário presente na erva-mate (Negrin et al., 2019). Induz efeitos estimulantes, que reduzem a fadiga, aumentam o estado de alerta e diminuem o tempo de reação. Quando consumida com moderação, produz efeitos positivos, como a redução do risco de fibrose hepática, câncer e depressão, mas, quando consumida em altas doses, pode levar a efeitos negativos, como insônia, nervosismo e disforia (van Dam et al., 2020). Assim, conhecendo a combinação de fatores ideal que promova a produção de plantas com maior controle sobre a composição dos metabólitos de interesse, é possível se inserir em um mercado crescente que atende consumidores com restrição a substâncias específicas.

O interesse na aplicação e desenvolvimento de técnicas para melhoramento genético em organismos como plantas vêm crescendo paralelo ao avanço da tecnologia. Programas conduzidos pela Embrapa Florestas no Brasil relataram a seleção de genótipos de erva-mate com características adaptativas desejáveis para a produção de clones objetivando alta produtividade e tolerância a estresses ambientais (Wendling et al., 2018). Estudos sobre o comportamento de clones de erva-mate submetidos a diferentes condições ambientais permitiram a produção de folhas com variados conteúdos de cafeína (de Almeida et al., 2022). Variações nos

teores de proteínas, compostos fenólicos e metilxantinas foram encontradas em folhas de erva-mate submetidas a condições de sombreamento e adubação diferenciadas. Tais resultados contribuem para a tomada de decisão em setores das indústrias alimentícias e de insumos medicinais (Westphalen et al., 2022)

Se tratando de organismos modificados quimicamente, análises constantes destes materiais devem ser realizadas de modo a garantir a qualidade e a estabilidade do processo de modificação. Neste sentido, métodos instrumentais de análise direta tem se destacado devido resultados promissores em pesquisas recentes (Silva et al., 2021; Tormena et al., 2020). A Espectroscopia no Infravermelho Próximo (do inglês, Near infrared, NIR) tem sido empregada por se tratar de uma técnica não destrutiva que fornece resultados rápidos e requer pouco ou nenhum pré-tratamento da amostra (Pasquini, 2003). Aliado a tratamentos quimiométricos adequados, tanto para discriminação, como a PLS-DA (do inglês, Partial least squares – Discriminant Analysis), quanto para a exploração, como a PCA (do inglês, Principal Component Analysis) e o ASCA (do inglês, ANOVA-simultaneous component analysis), é possível extrair a maior quantidade possível de informações presentes nos dados (Marcheafave, Pauli, et al., 2020; Tormena et al., 2019).

Este estudo tem o objetivo de diferenciar amostras de erva-mate a partir da aplicação de métodos quimiométricos a espectros NIR obtidos por análise direta do material vegetal visando o monitoramento de plantas quimicamente diferenciadas e submetidas a diferentes níveis de incidência luminosa. Verificar a aplicabilidade das metodologias propostas, em análise de campo ou como alternativa a técnicas analíticas que possuem maior custo relativo além de gerar resíduos tóxicos, por meio de modelos PLS-DA e análise multivariada ASCA.

Referências

- de Almeida, A. G., Pauli, E. D., Tormena, C. D., Wendling, I., Rakocevic, M., Bruns, R. E., Scarminio, I. S., & Marcheafave, G. G. (2022). Portable NIR Spectroscopy-Chemometric Identification of Chemically Differentiated Yerba Mate (*Ilex paraguariensis*) Clones. *Food Analytical Methods*, 0123456789. <https://doi.org/10.1007/s12161-022-02431-y>
- Fayad, E., El-Sawalhi, S., Azizi, L., Beyrouthy, M., & Abdel-Massih, R. M. (2020). Yerba Mate (*Ilex paraguariensis*) a potential food antibacterial agent and combination assays with different classes of antibiotics. *Lwt*, 125(March), 109267. <https://doi.org/10.1016/j.lwt.2020.109267>
- Marcheafave, G. G., Pauli, E. D., Tormena, C. D., Mattos, L. E., de Almeida, A. G., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2020). Irrigated and CO₂ level effects on metabolism in *Coffea arabica* beans from mixture design – near infrared fingerprints. *Microchemical Journal*, 152(June 2019), 104276. <https://doi.org/10.1016/j.microc.2019.104276>
- Mazur, L., Peralta-Zamora, P. G., Demczuk, B., & Ribani, R. H. (2014). Application of multivariate calibration and NIR spectroscopy for the quantification of methylxanthines in yerba mate (*Ilex paraguariensis*). *Journal of Food Composition and Analysis*, 35(2), 55–60. <https://doi.org/10.1016/j.jfca.2014.04.005>
- Negrin, A., Long, C., Motley, T. J., & Kennelly, E. J. (2019). LC-MS Metabolomics and Chemotaxonomy of Caffeine-Containing Holly (*Ilex*) Species and Related Taxa in the Aquifoliaceae. *Journal of Agricultural and Food Chemistry*, 67(19), 5687–5699. <https://doi.org/10.1021/acs.jafc.8b07168>
- Oh, J., Jo, H., Cho, A. R., Kim, S. J., & Han, J. (2013). Antioxidant and antimicrobial activities of various leafy herbal teas. *Food Control*, 31(2), 403–409. <https://doi.org/10.1016/j.foodcont.2012.10.021>
- Pasquini, C. (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14(2), 198–219. <https://doi.org/10.1590/S0103-50532003000200006>
- Pauli, E. D., Scheel, G. L., Delaroza, F., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2019). Photodiode array chromatographic-spectrophotometric metabolite quantification for yerba-mate plant sexual dimorphism differentiation. *Microchemical Journal*, 151. <https://doi.org/10.1016/j.microc.2019.104218>
- Rakocevic, M., Costes, E., & Assad, E. D. (2011). Structural and physiological sexual dimorphism estimated from three-dimensional virtual trees of yerba-mate (*Ilex paraguariensis*) is modified by cultivation environment. *Annals of Applied Biology*, 159(2), 178–191. <https://doi.org/10.1111/j.1744-7348.2011.00484.x>
- Rakocevic, M., & Martim, S. F. (2011). Time series in analysis of yerba-mate biennial growth modified by environment. *International Journal of Biometeorology*, 55(2), 161–171. <https://doi.org/10.1007/s00484-010-0322-4>
- 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. (2021). 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(July 2020). <https://doi.org/10.1016/j.jfca.2020.103795>
- Takeshi Matsunaga, F., Rakocevic, M., & Brancher, J. D. (2014). Modeling the 3D structure and rhythmic growth responses to environment in dioecious yerba-

- mate. *Ecological Modelling*, 290(C), 34–44.
<https://doi.org/10.1016/j.ecolmodel.2013.10.035>
- 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*, 26(29), 30356–30364.
<https://doi.org/10.1007/s11356-019-06163-1>
- Tormena, C. D., Pauli, E. D., Marcheafave, G. G., Scheel, G. L., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2020). FT-IR biomarkers of sexual dimorphism in yerba-mate plants: Seasonal and light accessibility effects. *Microchemical Journal*, 158(July), 105329. <https://doi.org/10.1016/j.microc.2020.105329>
- van Dam, R. M., Hu, F. B., & Willett, W. C. (2020). Coffee, Caffeine, and Health. *New England Journal of Medicine*, 383(4), 369–378.
<https://doi.org/10.1056/nejmra1816604>
- Wendling, I., Sturion, J. A., Stuepp, C. A., Reis, C. A. F., Ramalho, M. A. P., & de Resende, M. D. V. (2018). Early selection and classification of yerba mate progenies. *Pesquisa Agropecuaria Brasileira*, 53(3), 279–286.
<https://doi.org/10.1590/S0100-204X2018000300002>
- Westphalen, D. J., Angelo, A. C., Rossa, Ü. B., Helm, C. V., Radetski, C. M., & Gomes, E. N. (2022). *Phytochemical composition of yerba mate leaves (Ilex paraguariensis) and its relation with cultivation conditions*. 2020, 99–107.

Manuscrito publicado na revista *Microchemical Journal*, v. 191, p. 108828, 2023.

**Determinação Espectral NIR Direta de Melhoramento Genético,
Disponibilidade Luminosa e seus efeitos de interação na
Composição Química de Folhas de Erva-Mate Quimicamente
Selecionadas**

*Andressa Gomes de Almeida^a, Cláudia Domiciano Tormena^a, Natália Saudade de
Aguilar^b, Ivar Wendling^c, Miroslava Rakocevic^c, Elis Daiane Pauli^d, Ieda Spacino
Scarmínio^a, Roy Edward Bruns^{*d}, Gustavo Galo Marcheafave^{*d}*

^a*Laboratório de Quimiometria em Ciências Naturais (LQCN), Departamento de
Química, Universidade Estadual de Londrina, P.O.Box 6001, 86051-990 Londrina-
PR, Brasil.*

^b*Departamento de Ciências Florestais, Universidade Federal do Paraná, 80210-170
Curitiba-PR, Brasil.*

^c*Embrapa Florestas, P.O.Box 319, 83411-000 Colombo-PR, Brasil.*

^d*Instituto de Química, Universidade Estadual de Campinas, P.O.Box 6154, 13083-
970 Campinas-SP, Brasil.*

*bruns@unicamp.br (R.E. Bruns), ggalo@unicamp.br (G.G. Marcheafave)

Resumo

O monitoramento das modificações químicas de plantas fornece informações úteis sobre a variabilidade genética em decorrência a estresses ambientais. Isso permite determinar se o padrão de compostos químicos de interesse permanece constante. Tradicionalmente, os compostos químicos são monitorados por meio de procedimentos analíticos clássicos que são caros e prejudiciais ao meio ambiente. Árvores de erva-mate descafeinadas e com alto teor de cafeína foram selecionadas com base em seus perfis químicos. Seu comportamento foi avaliado em cinco diferentes níveis de sombreamento (0%, 40%, 51%, 76% e 82%), imitando as características ambientais da erva-mate em sistemas agroflorestais. Impressões digitais por espectroscopia de infravermelho próximo (NIR) foram obtidas para folhas de erva-mate e modelos de análise de variância - análise de componentes simultâneos (do inglês, ANOVA – Simultaneous Component Analysis, ASCA) foram desenvolvidos para monitoramento químico. O pré-processamento dos espectros NIR foi avaliado através de um planejamento fatorial completo de três níveis. O melhor conjunto de condições de pré-processamento encontrado foi com a variável normal padronizada (do inglês, standard normal variate, SNV) seguida por segundas derivadas, calculadas de acordo com o método Savitzky–Golay (janela de 11 pontos e polinômio interpolador de segundo grau). Os modelos ASCA mostraram satisfatoriamente diferenças significativas no comportamento químico para o fator do melhoramento genético, disponibilidade de luz e o efeito de interação entre esses fatores principais. O efeito do melhoramento genético teve uma variância de 44,30% sobre as impressões digitais NIR, indicando que as diferenças químicas são mantidas em qualquer nível de luz. O efeito de interação com 1,84% da variância total ilustrou que os comportamentos da erva-mate descafeinada e com alto teor de cafeína foram diferentes quando submetidos ao mesmo nível de luz. As principais informações indicadas pelo modelo ASCA foram o estiramento O-H e a banda combinada de flexão em 1654 nm, e o estiramento –CH do 1º sobretom aromático em 1675 nm. O método ASCA ofereceu uma ferramenta poderosa para monitorar plantas quimicamente selecionadas combinadas com espectroscopia NIR. O método é de baixo custo, rápido, ecológico e pode ser utilizado para pesquisas de campo.

Palavras-chave: *Ilex paraguariensis*; descafeinado; alto teor de cafeína; modificação química; ASCA.

1. Introdução

Com o aumento da informação a respeito das propriedades benéficas e das aplicações da erva-mate (*Ilex paraguariensis*), o interesse em compreender sua química se intensifica. Esta espécie de planta, catalogada em meados do século XIX por Auguste de Saint-Hilaire (Croge et al., 2020), era historicamente utilizada por povos indígenas Guarani. Após essa exploração pós-colombiana inicial pelos Jesuítas, por volta do século XVI, esta planta começou a ser difundida para os outros continentes (Gawron-gzella & Chanaj-kaczmarek, 2021). A erva-mate pertence a família das Aquifoliaceae, e ocorre naturalmente em florestas de Araucárias (Jr et al., 2002) encontradas no continente sul-americano, no norte da Argentina, sul do Brasil, Paraguai, e Uruguai (Gawron-gzella & Chanaj-kaczmarek, 2021). Alguns efeitos positivos de seu consumo podem ser destacados, como: atividade antimicrobiana, antioxidante, antiobesidade, e propriedades antidiabetes, além de melhoras no sistema digestivo e cardiovascular (Burris et al., 2012). Estas propriedades relacionadas a sua composição química, rica em compostos bioativos, como os alcalóides, polifenóis, saponinas, óleos essenciais, vitaminas e minerais, são as principais responsáveis por atrair os interesses da ciência e da indústria mundial (Croge et al., 2020).

O consumo de altas concentrações de substâncias como a cafeína, presente naturalmente na erva-mate, tem sido associado a efeitos adversos, como o aumento da pressão sanguínea, insônia, ansiedade, nervosismo, e disforia. Em contrapartida, pode proporcionar efeitos benéficos na redução da fadiga, aumentar o estado de alerta, reduzir o apetite e até mesmo reduzir dores quando associado a agentes analgésicos, entre outras vantagens (van Dam et al., 2020). Neste sentido, há algumas décadas, estudos a respeito da produção de plantas quimicamente modificadas têm sido desenvolvidos com o objetivo de controlar seu conteúdo químico, aumentando a concentração de algum metabólito de interesse ou eliminando-o, como, por exemplo, a cafeína (Wendling, 2016).

A nível estrutural, regulações epigenéticas, que não alteram o DNA da planta mas alteram sua expressão, podem ocorrer por meio de três vias: metilação do DNA, modificação de histona e interferência de RNA, em decorrência a mudanças ambientais - efeitos bióticos e abióticos (Singroha & Sharma, 2019). Programas conduzidos pela Embrapa Florestas no Brasil, reletaram a seleção de genótipos de

erva-mate com características adaptativas desejáveis, as quais tem se correlacionado a alterações no conteúdo de substâncias químicas relevantes, para a produção de clones objetivando alta produtividade e tolerância a estresses ambientais (Wendling et al., 2018).

Estudos sobre o comportamento de clones de erva-mate submetidos a diferentes condições ambientais, permitiram a produção de folhas com variados conteúdos de cafeína (de Almeida et al., 2022). Plantas idênticas geneticamente (clones) foram analisadas em termos da produtividade de biomassa, onde foi possível verificar a influência da exposição a diferentes níveis de sombreamento (de Aguiar et al., 2022). Tais estudos pretendem avaliar a combinação ideal de fatores que possam atender a um objetivo pré-especificado, uma vez que estas espécies são cultivadas em diferentes sistemas de plantio, como o agroflorestal (cultivo em bosque, sombreado, com menor exposição a luz solar), e a monocultura (promove maior exposição a luz do sol) (de Aguiar et al., 2022; Gawron-gzella & Chanaj-kaczmarek, 2021; Rakocevic et al., 2023).

Em geral, a composição química das plantas pode ser afetada pelo genótipo, sazonalidade, ciclo circadiano, idade e desenvolvimento das folhas, bem como pelo clima, disponibilidade de água, intensidade de luz, adição de nutrientes, altitude, poluição atmosférica e danos físicos (Gobbo-Neto & Lopes, 2007; Rakocevic et al., 2012). Alguns estudos indicam variações no teor de metabólitos secundários da erva-mate devido às condições ambientais, como clima, variabilidade genética, tipo de cultivo, idade da planta e condições agrônômicas (Pauli et al., 2019; Tormena et al., 2020). Variações nos teores de proteínas, compostos fenólicos e metilxantinas foram encontradas em folhas de erva-mate submetidas a diferentes condições de sombreamento e adubação. Tais resultados favorecem os interesses das indústrias alimentícias e de insumos medicinais (Westphalen et al., 2022). O efeito do estresse ambiental sobre o metaboloma da erva-mate torna-se um desafio na seleção de plantas quimicamente modificadas que devem ter independência metabólica quanto aos efeitos abióticos, ou seja, sua particularidade química selecionada deve ser mantida independente do ambiente, confirmando um organismo geneticamente melhorado.

Devido ao rigor no controle de substâncias químicas nos insumos alimentícios, como um 'termômetro' avaliativo da qualidade e segurança alimentar, são necessárias técnicas de análise sofisticadas e instrumentos que forneçam resultados

de alta precisão mesmo para baixas concentrações de analito, como cromatografia, espectrofotometria UV-VIS, espectrometria de massa, fluorescência e luminescência. Geralmente, a utilização dessas técnicas está condicionada ao pré-tratamento das amostras e requer grandes volumes de solvente (Pallone et al., 2018). Nesse sentido, a busca por novas metodologias capazes de solucionar este problema torna-se necessária. Assim, o uso da espectroscopia vibracional, como o infravermelho próximo (NIR), atende aos princípios da 'química verde', sendo uma alternativa que rapidamente fornece dados confiáveis a um baixo custo relativo, utilizando pouco ou nenhum solvente que poderia ser tóxico para o ambiente (Beć et al., 2021; Marcheafave, Pauli, et al., 2020; Pallone et al., 2018; Tormena et al., 2019).

A espectroscopia NIR opera na região entre 800 – 2500 nm, caracteriza-se por observar “transições proibidas”, ou seja, harmônicos e bandas de combinação. Esta característica permite que a técnica analise diferentes grupos de moléculas (Beć et al., 2021). No entanto, o tratamento quimiométrico é essencial para adquirir informações úteis do NIR, pois a leitura das amostras gera uma grande quantidade de dados com bandas sobrepostas que requerem processamento para posterior interpretação (Pallone et al., 2018). Entre as ferramentas estatísticas utilizadas, a análise de variância - análise de componentes simultâneos (do inglês, ANOVA - Simultaneous Component Analysis) se adequa perfeitamente quando o objetivo é analisar as diferentes contribuições das variantes causadas por fatores relacionados a um planejamento experimental ao qual as amostras foram submetidas (Cheng et al., 2019). Essa ferramenta multivariada fornece a base para a interpretação de conjuntos de dados complexos e tem sido amplamente explorada em matrizes de alimentos para analisar efeitos abióticos, avaliar parâmetros de qualidade e detectar fraudes (Rust et al., 2021). Nossa hipótese é que modelos ASCA construídos com impressões digitais NIR adquiridas diretamente em folhas de erva-mate quimicamente modificadas sob estresse ambiental forneceriam informações valiosas sobre o monitoramento de mudanças químicas e garantiriam a seleção correta de plantas melhoradas geneticamente.

O presente trabalho tem como objetivo avaliar os efeitos do melhoramento genético, disponibilidade de luz e suas interações em folhas de erva-mate com alto teor de cafeína e descafeinada por meio de análise multivariada ASCA e dados fornecidos por espectroscopia NIR, para verificar a aplicabilidade dessas

ferramentas ecologicamente corretas como metodologia para monitorar plantas quimicamente selecionadas submetidas a condições de estresse ambiental, em oposição a vários métodos analíticos agressivos ao meio ambiente.

2. Materiais e Métodos

2.1 *Produção da planta e amostragem*

Plantas de erva-mate foram cultivadas em Colombo, Paraná, Brasil (25°19'15" S, 49°09'31" W; 934 m) em um sistema de cultivo semi-hidropônico com areia de grão médio (camadas de areia) em calhas. As plantas foram cultivadas em casa de vegetação plástica (média anual de 28 °C). Os genótipos clonais foram propagados por miniestacas, plantadas em espaçamento de 15 cm x 15 cm e irrigação por gotejamento com a solução nutritiva, conforme de Aguiar et al., 2022. Os genótipos de erva-mate aqui utilizados fazem parte do acervo de melhoramento genético da Embrapa Florestas, Colombo, Brasil.

Os genótipos EC22 e EC40 foram selecionados com base nos níveis de compostos bioativos (de Aguiar et al., 2022; Sturion et al., 2017; Vieira Helm et al., 2015; Wendling, 2016; Wendling et al., 2007, 2018). O genótipo EC40 possui alto teor de cafeína (2,35% de cafeína) nas folhas, e o genótipo EC22 é descafeinado (0,02% de cafeína); para esta análise, folhas maduras foram coletadas de plantas adultas cultivadas a pleno sol (Vieira Helm et al., 2015). A categorização da cafeína seguiu Schuhli et al., 2019: alto teor de cafeína significa > 1,6% (16 mg g⁻¹) e descafeinado < 0,1% (1 mg g⁻¹). Os níveis de sombreamento foram obtidos pela cobertura com telas de polietileno comercial sobre as plantas (0%, 40%, 51%, 76% e 82% de sombreamento). Os clones selecionados desenvolveram-se no sistema descrito acima, e as folhas foram coletadas em maio de 2021. Após a colheita, as folhas foram secas em forno de micro-ondas (Tomasi et al., 2021) e armazenadas em ambiente com temperatura controlada (20 °C).

2.2 *Medições NIR*

As medições NIR foram realizadas de acordo com de Almeida et al., 2022. As medições de refletância foram realizadas diretamente no material vegetal em dez posições diferentes de uma lâmina foliar usando o software DLP NIRscan Nano (Texas Instruments Inc., Dallas, TX, EUA) contra um branco a 25 ± 1 °C. Cada

espectro foi obtido pela média de 99 varreduras entre 900 e 1700 nm com resolução de 3,5 nm.

2.3 *Tratamentos quimiométricos*

2.3.1 *Efeitos de pré-processamento em resultados ASCA*

Em quimiometria, o pré-processamento de NIR é uma etapa crucial para a obtenção de modelos robustos que contenham informações químicas relevantes para classificação, regressão ou análise exploratória por meio da eliminação ou suavização de fenômenos físicos que interferem e atrapalham o uso de espectros na região NIR (Pasquini, 2003). O pré-processamento visa reduzir a variância residual dos dados de modelagem para otimizar o perfil espectral do objetivo proposto (Rinnan et al., 2009). Embora amplamente discutido na literatura, a maioria das escolhas de pré-processamento e suas combinações são realizadas de forma aparentemente arbitrária e aleatória. Nos casos de análise exploratória, como a análise de componentes principais, a interpretação dos resultados é feita qualitativamente, mas em modelos com resultados quantitativos é possível avaliar estatisticamente a eficiência de diferentes combinações de técnicas de pré-processamento. Neste trabalho, propomos avaliar o uso e as combinações de pré-processamento por meio do planejamento de experimentos. Como triagem experimental, foi usado um planejamento fatorial completo de três níveis (3^3) (BENNETT, 2006). Os fatores selecionados foram: fatores de normalização, derivação e média móvel. Para o fator de normalização, os níveis foram: variável normal padronizada (SNV, do inglês, standard normal variate), sem normalização, e correção de espalhamento multiplicativo (MSC, do inglês, multiplicative signal correction). Para o fator de derivação, os níveis foram 1ª derivada, sem derivação e 2ª derivada. Para o fator de média móvel, os níveis aplicados foram 3, 7 e 11 janelas. A representação e codificação do delineamento experimental são apresentadas na Figura 1 e na Tabela 1. Os níveis designados como "sem" ou "0" são essenciais para verificar a individualidade de cada pré-processamento nos resultados da modelagem ASCA. MSC e SNV pertencem ao chamado grupo de métodos de correção de dispersão (Rinnan et al., 2009). As derivadas espectrais aqui são empregadas como filtros de derivação polinomial Savitzky-Golay (SG) (1ª e 2ª derivadas com polinômio interpolador de segundo grau) (Rinnan et al., 2009). A

média móvel suaviza o espectro antes das derivações, a fim de diminuir o efeito prejudicial na relação sinal-ruído (Rinnan et al., 2009).

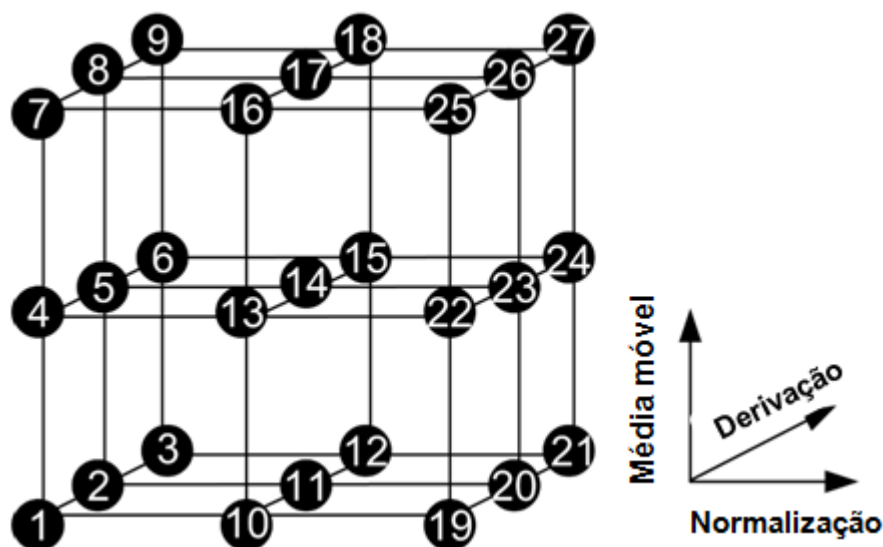


Fig 1. Planejamento experimental realizado como triagem de pré-processamento para modelagem ASCA com planejamento fatorial completo de três níveis (3^3) com os fatores normalização, derivação e média móvel.

2.3.2 Modelos ASCA

A experimentação em metabolômica avançou nas últimas décadas com o auxílio do planejamento de experimentos (DoE). Assim, a complexidade dos fatores e suas respostas atingiram níveis que exigiram modelos multivariados para elucidar os resultados (Jansen et al., 2005). Conjuntos de dados contínuos, como cromatogramas ou espectros moleculares derivados de DoE, continham respostas químicas de interesse, mas precisavam ser resolvidos para elucidar a importância de fatores experimentais na resposta química (Smilde et al., 2005). Os principais efeitos e interações entre os fatores experimentais permitem determinar quais variáveis são importantes na alteração química e o quanto elas interferem no perfil químico em análise. Essa complexidade foi resolvida satisfatoriamente com a análise de componentes simultâneos ANOVA (ASCA) (Zwanenburg et al., 2011). O método combina o aspecto quantitativo da ANOVA para determinar valores de efeito significativos, juntamente com as exibições gráficas da análise de componentes simultâneos (SCA), permitindo a interpretação de suas possíveis consequências nas composições químicas.

O ASCA tem se estabelecido continuamente na área biológica e química

através da metabolômica (Smilde et al., 2005), mas sua aplicabilidade traz inúmeras vantagens para a análise metabólica de dados da área agrônômica (Silva et al., 2021). Seu uso no campo avança através da codificação de fatores experimentais, tais como condições ambientais e espécies de plantas – questionando o panorama químico como resposta indutora dos fatores (Marcheafave, Tormena, et al., 2020). Em geral, o ASCA permite determinar se os fatores ambientais levam a alterações químicas significativas na planta em estudo e se existem interações entre os diferentes sistemas aos quais ela está sujeita. A descrição do modelo ASCA, seu desenvolvimento e aplicações podem ser encontradas em Jansen et al., 2005, Smilde et al., 2005 e Timmerman et al., 2015.

Para as análises ASCA, a matriz codificada continha dois fatores (melhoramento genético e sombreamento). Os níveis de melhoramento genético foram clones descafeinados e com alto teor de cafeína, sendo 0%, 40%, 51%, 76% e 82% os níveis para o fator de sombreamento. Cada combinação de níveis entre fatores continha 200 medições espectrais NIR, resultando em uma matriz de 2.000 × 228. O número de linhas (2000) representava diferentes espectros NIR para as diferentes condições, enquanto as colunas (228) descreviam os valores de refletância NIR entre 900 e 1700 nm.

Todos os modelos foram processados usando o PLS toolbox 8.7.1[®] da Eigenvector Research com o software Matlab 2016[®].

3. Resultados e Discussão

O método estatístico ASCA foi aplicado aos dados espectrais NIR brutos das folhas de erva-mate (Fig S1) para determinar a significância dos fatores do planejamento experimental (melhoramento genético, sombreamento e suas interações). Tendo em vista que alguns efeitos instrumentais podem prejudicar ou piorar o desempenho dos modelos quimiométricos, diferentes tipos de pré-tratamentos espectrais foram testados. O planejamento fatorial completo para a exploração do pré-processamento, sua codificação e resultados são mostrados na Tabela 1. Embora o planejamento de três níveis seja oneroso em termos de número de execuções, neste caso, ele exige apenas a execução matemática realizada por software, o que torna o processo simples, rápido, muito mais completo e robusto do que execuções aleatórias para escolha de pré-processamento, como ocorre em

praticamente todos os casos disponíveis na literatura.

Tabela 1. Seleção do modelo por análise de componentes simultâneos-ANOVA (ASCA): pré-processamento espectral NIR na determinação de valores de efeito para o melhoramento genético, sombreamento e suas interações em folhas de erva-mates quimicamente modificadas.

Fatores	-	0	+
1. Normalização SNV		Sem	MSC
2. Média móvel	3	7	11
3. Derivação	1 ^a	Sem	2 ^a

Experimento fatorial	Nível de fator 1 2 3	Resultados ASCA			
		Efeito do melhoramento genético (%)	Efeito do sombreamento (%)	Efeito da interação (%)	Residuais (%)
1	---	28.53	4.27	3.05	64.15
2	--0	5.79	4.75	3.52	85.94
3	--+	3.64	2.57	3.08	90.71
4	-0-	28.48	4.43	2.58	64.51
5	-00	4.45	4.73	3.53	87.28
6	-0+	28.14	3.86	1.74	66.26
7	-+-	28.45	4.16	2.72	64.67
8	-+0	2.69	4.72	3.60	88.98
9	-++	44.30	3.84	1.84	50.03
10	0--	20.41	3.62	7.08	68.89
11	0-0	4.94	3.89	7.96	83.22
12	0-+	11.75	6.64	8.19	73.42
13	00-	19.34	3.64	7.07	69.96
14	000	4.94	3.89	7.96	83.22
15	00+	30.65	5.83	4.57	58.95
16	0+-	16.54	2.98	7.63	72.85
17	0+0	4.87	3.89	7.97	83.27
18	0++	29.40	3.73	4.99	61.89
19	+--	7.33	1.25	0.81	90.56
20	+ -0	0.54	1.05	0.41	98.00
21	+ -+	0.49	0.55	0.48	98.48
22	+ 0-	7.85	1.47	0.72	89.97
23	+ 00	0.45	1.04	0.41	98.10
24	+ 0+	8.17	1.34	0.56	89.93
25	+ +-	7.24	1.38	0.81	90.56
26	+ +0	0.33	1.03	0.40	98.23
27	+ ++	22.49	1.57	0.91	75.03

A partir da Tabela 1 é possível observar a variância dos efeitos

principais e de interação e a variância residual para os respectivos modelos construídos após as diferentes combinações de pré-processamentos. Todos esses modelos foram determinados com dados espectrais centrados na média. Todos os resultados ASCA nesta tabela são significativos no nível de confiança de 95%, p-valor < 0,05, considerando 10.000 permutações. A característica mais marcante manifestada nesta tabela é como a sensibilidade dos valores de efeito dependem da combinação de métodos de pré-processamento usados para tratar os dados. Mesmo que os modelos ASCA possam ter uma grande porção de variância residual, efeitos significativos podem ser detectados quando a análise é realizada nas matrizes de efeito, em vez de diretamente nos dados espectrais. Infelizmente, não há garantia de que todos os modelos ASCA serão significativos, independentemente das técnicas de pré-processamento, como descobrimos no tratamento de outros sistemas de dados.

Assumimos que quanto maior as variâncias explicadas para os efeitos ASCA, modelos mais completos foram alcançados; assim, os valores de efeito para o ASCA mostraram que a normalização apresenta valores maiores no nível negativo (SNV) do que os obtidos no nível positivo (MSC). Quando o fator de normalização não é aplicado (nível 0), alguns resultados satisfatórios são encontrados, mas ainda são inferiores quando comparados com o SNV (Tabela 1). Assim, SNV correspondeu a melhores resultados na normalização de impressões digitais NIR do que MSC ou sem normalização. Para a média móvel, a variação dos efeitos é levemente pronunciada quando associada a SNV ou sem normalização. Por outro lado, menores porcentagens de variância para os modelos ASCA são observados quando usados concomitantemente com MSC. A diferença na variância provavelmente se deve a interações entre a normalização e os fatores de média móvel. Em geral, a média móvel de 11 janelas mostra uma variação mais significativa dos efeitos ASCA. Por sua vez, o fator de derivação não causa variações substanciais na modelagem ASCA em relação aos demais fatores aqui estudados.

Outra observação crítica na escolha do modelo ASCA envolve a variância atribuída aos resíduos. O SNV seguido pela segunda derivada, calculado de acordo com o método Savitzky-Golay (janela de 11 pontos e polinômio interpolador de segundo grau) (Fig S2) e, finalmente, a centralização média levou aos melhores resultados em termos de menor valor residual (Tabela 1). Neste caso, cerca de 50% da variância total deveu-se à variação residual, ou seja, a variação não representada

pelo modelo e que correspondeu às diferenças espectrais do sinal entre as réplicas. Assim, o modelo 9 na Tabela 1 apresenta uma escolha racional para o pré-tratamento desses dados NIR e interpretação dos resultados ASCA.

O melhoramento genético tem o maior efeito na variabilidade espectral NIR, sendo responsável por mais de 44% da variação total de acordo com este modelo. A matriz correspondente a este efeito foi posteriormente investigada usando SCA. Como o melhoramento genético foi estudado em dois níveis, apenas um componente pode ser analisado. A Figura 2 mostra a variação entre as réplicas no gráfico de escores da PC uma vez que os resíduos são projetados no subespaço da componente principal da matriz de efeito de melhoramento genético. Os escores das folhas de erva-mate com alto teor de cafeína se localizaram em sua maioria na região positiva, enquanto a grande maioria das descafeinadas se projetou na porção negativa deste gráfico (Figura 2).

Os valores de loadings mostraram que a região espectral mais afetada pela variação no melhoramento genético foi aquela entre 1600 e 1700 nm (Figura 3). O comprimento de onda na região de 1654 nm tem loadings positivos, correspondendo às amostras com alto teor de cafeína, superiores a 1,6% de cafeína. As amostras descafeinadas com menos de 0,1% de cafeína tiveram loadings negativos em 1675 nm. As principais absorções na região NIR foram obtidas através de harmônicos e combinações de vibrações fundamentais de $-CH$, $-NH$, $-OH$ (Türker-Kaya & Huck, 2017). O estiramento O-H e a banda combinada de flexão O-H podem ser atribuídos ao comprimento de onda em 1654 nm, enquanto o estiramento $-CH$ do 1º harmônico aromático pode ser associado ao comprimento de onda em 1675 nm (Workman & Weyer, 2007).

A significância estatística do fator de melhoramento genético das impressões digitais NIR denota que a variação química que ocorre entre a planta descafeinada e com alto teor de cafeína é mantida independentemente da condição de luz. Ou seja, mantém-se a variação genética inerente às plantas de erva-mate e assegura-se a descafeinação. Assim, as impressões digitais NIR e os modelos ASCA foram capazes de prever se o perfil químico era persistente na seleção da erva-mate ou modificado pelo estresse ambiental.

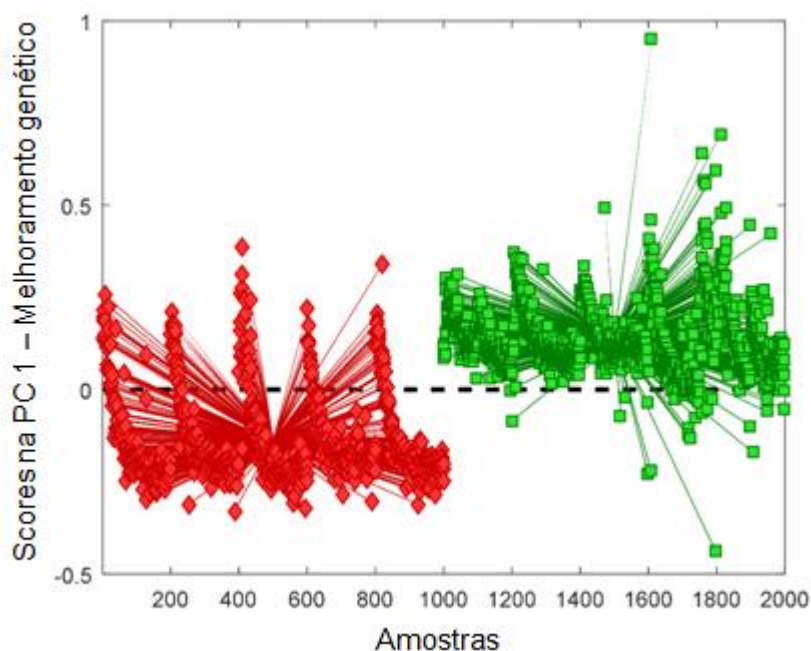


Fig 2. Gráfico de scores para o efeito do melhoramento genético em folhas de erva-mate (descafeinado (♦), ou com alto teor de cafeína (■)) com residuais projetados a partir de análise ASCA das impressões digitais NIR entre 900-1700 nm.

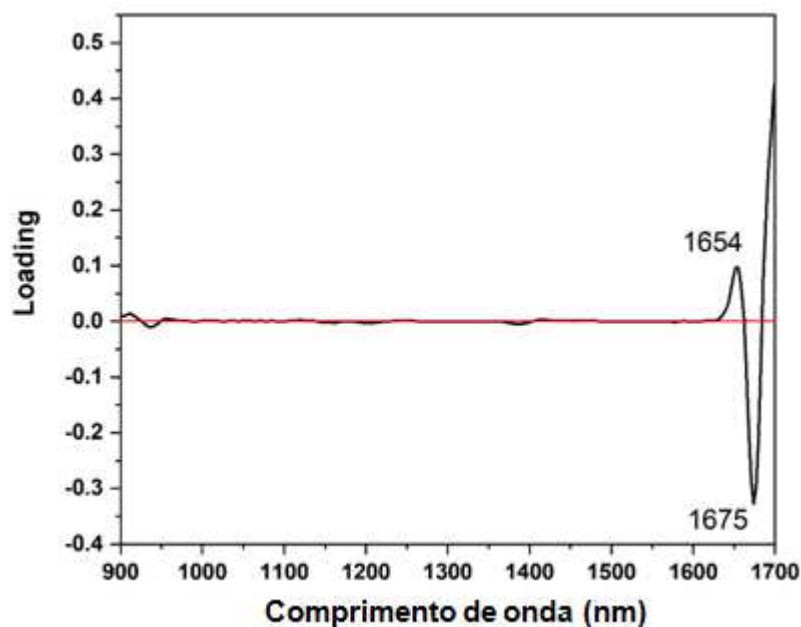


Fig 3. Gráfico de loading da SC1 para o efeito do melhoramento genético em folhas de erva-mate (alto teor de cafeína - 1654 nm, ou descafeinado - 1675 nm) a partir de análise ASCA das impressões digitais NIR entre 900-1700 nm.

Para o efeito do sombreamento, o modelo ASCA reteve 3,84% de variação dos

espectros NIR (Tabela 1). A projeção dos escores separou os níveis de sombreamento 0% e 40% que tendiam a ocorrer na região negativa, enquanto as amostras com nível de sombreamento 51% foram delimitadas em torno da posição zero na abscissa, enquanto aquelas com 76% e 82% de sombreamento foram projetadas mais para a região positiva (Figura 4). De acordo com os loadings, 1654 nm foi associado ao menor sombreamento (0 e 40%), enquanto 1675 nm foi associado ao maior sombreamento (76% e 82%) (Figura 5). O nível de irradiância é um fator importante no metabolismo secundário da erva-mate, e sua significância era esperada nas impressões digitais NIR, pois tem se mostrado um fator significativo na diferenciação química das folhas de erva-mate (Tormena et al., 2020). Nas folhas de erva-mate, os teores de cafeína e teobromina de plantas cultivadas sob alta intensidade de luz mostram grandes sensibilidades quando modificados pela época de colheita e estação do ano (Pauli et al., 2019).

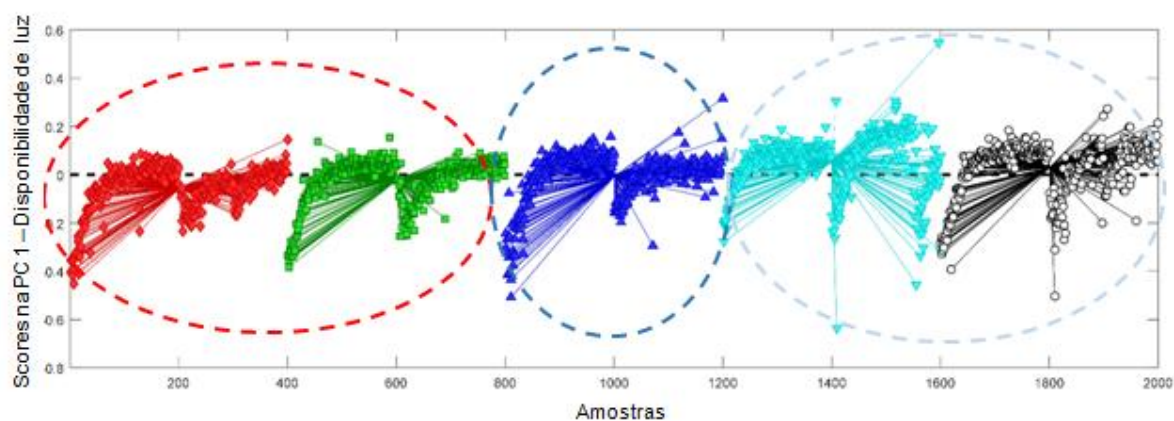


Fig. 4. Gráfico de escores para o efeito da disponibilidade de luz em folhas de erva-mate ((♦) 0%, (■) 40%, (▲) 51%, (▼) 76%, e (○) 82% níveis de sombreamento) com residuais projetados a partir de análise ASCA das impressões digitais NIR entre 900-1700 nm.

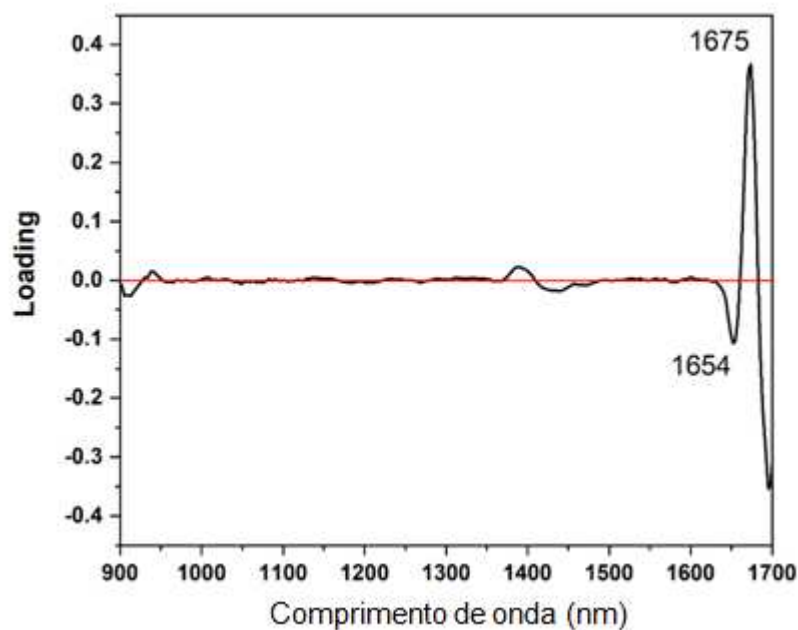


Fig. 5 Gráfico de loading da SC1 para o efeito da disponibilidade de luz em folhas de erva-mate a partir de análise ASCA das impressões digitais NIR entre 900-1700 nm.

Apesar de baixo (1,84%), o efeito da interação entre melhoramento genético e sombreamento foi significativo (Figura 6). De acordo com os loadings da primeira componente, a região espectral entre 1654 e 1675 nm contribuiu para a detecção dessa interação (Figura 7). A significância da interação demonstrou satisfatoriamente diferentes respostas químicas de plantas de erva-mate descafeinadas e com alto teor de cafeína expostas a diferentes níveis de sombreamento. A significância da interação entre melhoramento genético × disponibilidade de luz corroborou o fator 1, indicando que ambas as plantas apresentaram comportamentos genéticos distintos.

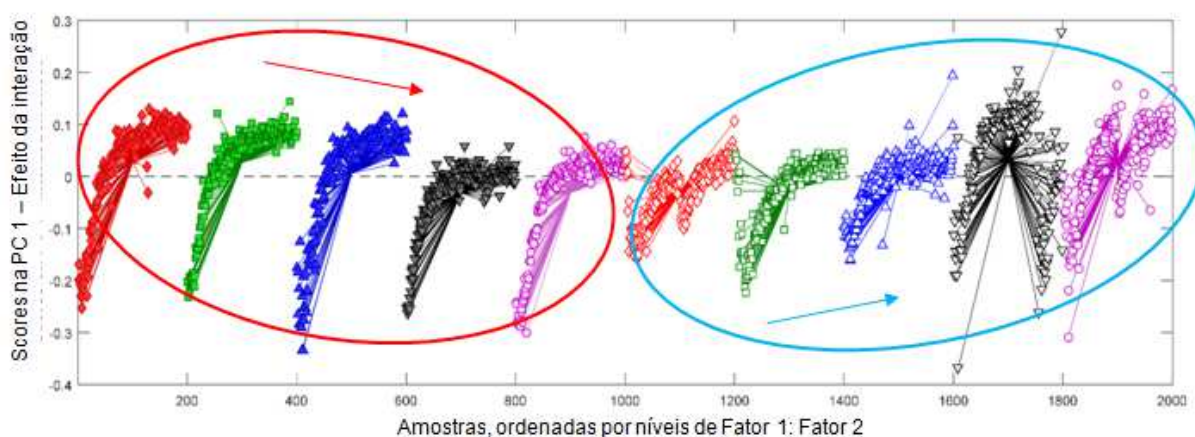


Fig. 6. Gráfico de escores para o efeito da interação entre o melhoramento genético e o sombreamento em folhas de erva-mate: ((♦) descafeinado sob 0% de sombreamento; (■) descafeinado sob 40% sombreamento; (▲) descafeinado sob 51% sombreamento; (▼) descafeinado sob 76% sombreamento; (●) descafeinado sob 82% sombreamento; (◇) alta cafeína sob 0% sombreamento; (□) alta cafeína sob 40% sombreamento; (△) alta cafeína sob 51% sombreamento; (▽) alta cafeína com 76% sombreamento; (○) alta cafeína com 82% sombreamento) com residuais projetados a partir de análise ASCA das impressões digitais NIR entre 900-1700 nm.

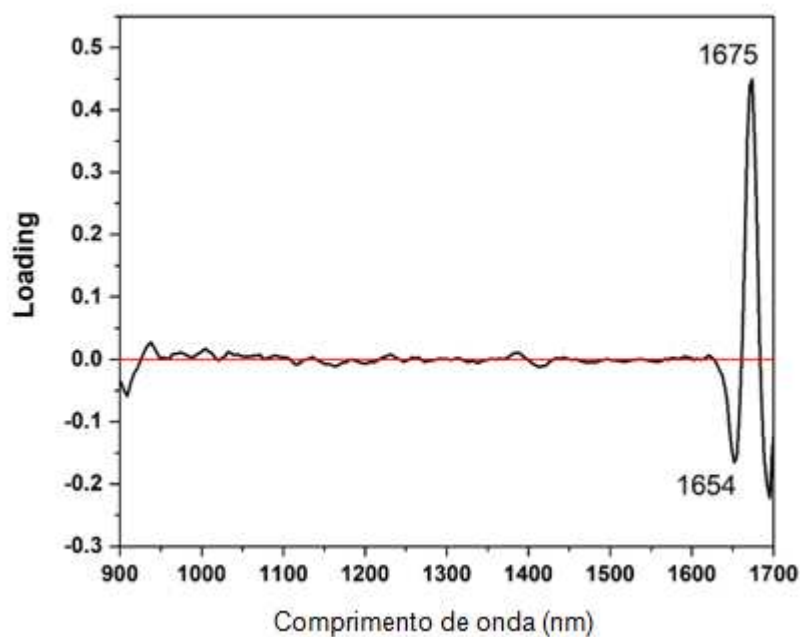


Fig. 7. Gráfico de loading da PC1 para o efeito da interação (melhoramento genético × disponibilidade de luz) em folhas de erva-mate a partir de análise ASCA das impressões digitais NIR entre 900-1700 nm.

Conclusão

O método ASCA associado a medições diretas de NIR em material vegetal permitiu o monitoramento de plantas de erva-mate quimicamente modificadas submetidas a vários níveis de sombreamento. Um planejamento fatorial completo permitiu uma investigação sistemática dos valores de variância ASCA devido aos diferentes métodos de pré-processamento empregados na modelagem NIR. O SNV seguido por segundas derivadas, calculado de acordo com o método Savitzky-Golay (janela de 11 pontos e polinômio interpolador de segundo grau) e, finalmente, centralização média foram os pré-tratamentos que levaram aos melhores resultados em termos de menor valor residual em modelos ASCA. Os resultados da ASCA mostraram que o efeito do melhoramento genético foi significativo independentemente da condição de luz - permaneceu constante em qualquer nível de sombreamento. A significância do fator de melhoramento genético indicou comportamentos distintos sob qualquer condição de luz. O sombreamento foi significativo no monitoramento das plantas de erva-mate quimicamente modificadas e foi responsável por cerca de 3,84% da variação total. Finalmente, a interação entre melhoramento genético e sombreamento confirmou a diferença genética entre as plantas modificadas quimicamente. Plantas de erva-mate descafeinadas e com alto teor de cafeína, ao receberem a mesma incidência solar, apresentam comportamentos químicos diferentes. NIR e ASCA podem fornecer análises rápidas, baratas e quimicamente amigáveis com resultados promissores no monitoramento da variação química de plantas genéticas selecionadas devido a atributos metabólicos de interesse.

Agradecimentos

ISS, REB, EDP, e MR agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelas bolsas de pesquisa (PQ, Processos: 302204/2018-0 e 302574/2019-0; PS, Processo: 102416/2022-2; PV, Processo: 350509/2020-4). GGM agradece pela concessão 2020/11463-5, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Referências

- Beć, K. B., Grabska, J., & Huck, C. W. (2021). Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers. *Chemistry - A European Journal*, 27(5), 1514–1532. <https://doi.org/10.1002/chem.202002838>
- BENNETT, G. (2006). R.E. Bruns, I.S. Scarminio and B. De Barros Neto, Statistical Design—Chemometrics, Elsevier B.V., Amsterdam, The Netherlands (2006) 422 pp., US\$ 145.00, ISBN: 0-444-52181-X. *Journal of Hazardous Materials*, 138(2), 417–417. <https://doi.org/10.1016/j.jhazmat.2006.05.113>
- Burris, K. P., Harte, F. M., Michael Davidson, P., Stewart, C. N., & Zivanovic, S. (2012). Composition and bioactive properties of Yerba Mate (*Ilex paraguariensis* A. St.-Hil.): A review. In *Chilean Journal of Agricultural Research* (Vol. 72, Issue 2). <https://doi.org/10.4067/s0718-58392012000200016>
- Cheng, W., Sørensen, K. M., Mongi, R. J., Ndabikunze, B. K., Chove, B. E., Sun, D. W., & Engelsens, S. B. (2019). A comparative study of mango solar drying methods by visible and near-infrared spectroscopy coupled with ANOVA-simultaneous component analysis (ASCA). In *Lwt* (Vol. 112). <https://doi.org/10.1016/j.lwt.2019.05.112>
- Croge, C. P., Cuquel, F. L., & Pintro, P. T. M. (2020). Yerba mate: Cultivation systems, processing and chemical composition. a review. *Scientia Agricola*, 78(5), 1–11. <https://doi.org/10.1590/1678-992x-2019-0259>
- de Aguiar, N. S., Moreno Gabira, M., de Cássia Tomasi, J., Mendes Duarte, M., Marcolino Vieira, L., José Lavoranti, O., & Wendling, I. (2022). Productivity of Clonal *Ilex paraguariensis* Genotypes in a Semi-Hydroponic System Is Reduced by Shading. *Forest Science*, 68(5–6), 540–547. <https://doi.org/10.1093/forsci/fxac028>
- de Almeida, A. G., Pauli, E. D., Tormena, C. D., Wendling, I., Rakocevic, M., Bruns, R. E., Scarminio, I. S., & Marcheafave, G. G. (2022). Portable NIR Spectroscopy-Chemometric Identification of Chemically Differentiated Yerba Mate (*Ilex paraguariensis*) Clones. *Food Analytical Methods*, 0123456789. <https://doi.org/10.1007/s12161-022-02431-y>
- Gawron-gzella, A., & Chanaj-kaczmarek, J. (2021). *Yerba Mate — A Long but Current History*.
- Gobbo-Neto, L., & Lopes, N. P. (2007). Plantas medicinais: Fatores de influência no conteúdo de metabólitos secundários. In *Química Nova* (Vol. 30, Issue 2, pp. 374–381). <https://doi.org/10.1590/S0100-40422007000200026>
- Jansen, J. J., Hoefsloot, H. C. J., Van Der Greef, J., Timmerman, M. E., Westerhuis, J. A., & Smilde, A. K. (2005). ASCA: Analysis of multivariate data obtained from an experimental design. *Journal of Chemometrics*, 19(9). <https://doi.org/10.1002/cem.952>
- Jr, M. G., Pirani, R., & Brasil, N. (2002). *Aquifoliaceae* 1. 2(1992), 31–37.
- Marcheafave, G. G., Pauli, E. D., Tormena, C. D., Mattos, L. E., de Almeida, A. G., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2020). Irrigated and CO2 level effects on metabolism in *Coffea arabica* beans from mixture design – near infrared fingerprints. *Microchemical Journal*, 152(June 2019), 104276. <https://doi.org/10.1016/j.microc.2019.104276>
- 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. <https://doi.org/10.1016/j.scitotenv.2020.142350>
- Pallone, J. A. L., Caramês, E. T. dos S., & Alamar, P. D. (2018). Green analytical chemistry applied in food analysis: alternative techniques. In *Current Opinion in Food Science* (Vol. 22). <https://doi.org/10.1016/j.cofs.2018.01.009>
- Pasquini, C. (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14(2), 198–219. <https://doi.org/10.1590/S0103-50532003000200006>
- Pauli, E. D., Scheel, G. L., Delarozza, F., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2019). Photodiode array chromatographic-spectrophotometric metabolite quantification for yerba-mate plant sexual dimorphism differentiation. *Microchemical Journal*, 151. <https://doi.org/10.1016/j.microc.2019.104218>
- Rakocevic, M., Janssens, M., & Schere, R. (2012). Light responses and gender issues in the domestication process of yerba-mate, a subtropical evergreen. In *Evergreens: Types, Ecology and Conservation* (Vol. 2008, Issue 10102, pp. 63–96).
- Rakocevic, M., Maia, A. D. H. N., Duarte, M. M., & Wendling, I. (2023). Secondary sexual dimorphism in biomass production of *Ilex paraguariensis* progenies associated with their provenances and morphotypes. *Experimental Agriculture*, 59. <https://doi.org/10.1017/S0014479722000552>
- Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. In *TrAC - Trends in Analytical Chemistry* (Vol. 28, Issue 10). <https://doi.org/10.1016/j.trac.2009.07.007>
- Rust, A., Marini, F., Allsopp, M., Williams, P. J., & Manley, M. (2021). Application of ANOVA-simultaneous component analysis to quantify and characterise effects of age, temperature, syrup adulteration and irradiation on near-infrared (NIR) spectral data of honey. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 253, 119546. <https://doi.org/10.1016/j.saa.2021.119546>
- Schuhli, G. S. e, Penteado Junior, J. F., & Wendling, I. (2019). *Descritores mínimos em cultivares de espécies florestais: uma contribuição para erva-mate*. 25.
- 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. (2021). 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(July 2020). <https://doi.org/10.1016/j.jfca.2020.103795>
- Singroha, G., & Sharma, P. (2019). Epigenetic Modifications in Plants under Abiotic Stress. In *Epigenetics*. <https://doi.org/10.5772/intechopen.84455>
- Smilde, A. K., Jansen, J. J., Hoefsloot, H. C. J., Lamers, R. J. A. N., van der Greef, J., & Timmerman, M. E. (2005). ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics*, 21(13). <https://doi.org/10.1093/bioinformatics/bti476>
- Sturion, J. A., Stuepp, C. A., & Wendling, I. (2017). Genetic parameters estimates and visual selection for leaves production in *Ilex paraguariensis*. *Bragantia*, 76(4), 492–500. <https://doi.org/10.1590/1678-4499.2016.419>
- Timmerman, M. E., Hoefsloot, H. C. J., Smilde, A. K., & Ceulemans, E. (2015). Scaling in ANOVA-simultaneous component analysis. *Metabolomics*, 11(5), 1265–1276. <https://doi.org/10.1007/s11306-015-0785-8>
- Tomasi, J. D. C., De Lima, G. G., Wendling, I., Helm, C. V., Hansel, F. A., De Godoy,

- R. C. B., Grunennvaldt, R. L., De Melo, T. O., Tomazzoli, M. M., & Deschamps, C. (2021). Effects of different drying methods on the chemical, nutritional and colour of yerba mate (*Ilex paraguariensis*) leaves. *International Journal of Food Engineering*, 17(7), 551–560. <https://doi.org/10.1515/ijfe-2020-0312>
- 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*, 26(29), 30356–30364. <https://doi.org/10.1007/s11356-019-06163-1>
- Tormena, C. D., Pauli, E. D., Marcheafave, G. G., Scheel, G. L., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2020). FT-IR biomarkers of sexual dimorphism in yerba-mate plants: Seasonal and light accessibility effects. *Microchemical Journal*, 158(July), 105329. <https://doi.org/10.1016/j.microc.2020.105329>
- Türker-Kaya, S., & Huck, C. W. (2017). A review of mid-infrared and near-infrared imaging: Principles, concepts and applications in plant tissue analysis. In *Molecules* (Vol. 22, Issue 1). <https://doi.org/10.3390/molecules22010168>
- van Dam, R. M., Hu, F. B., & Willett, W. C. (2020). Coffee, Caffeine, and Health. *New England Journal of Medicine*, 383(4), 369–378. <https://doi.org/10.1056/nejmra1816604>
- Vieira Helm, C., Ruiz, H. Z., Hansel, F. A., Stuepp, C. A., & Wendling, I. (2015). *Técnico Comunicado Efeito do solvente na extração de teobromina e cafeína em progênies de erva-mate*. 3–8.
- Wendling, I. (2016). Melhoramento de erva-mate: perspectivas. In: *Embrapa Florestas (Ed.), SEMINÁRIO ERVA-MATE XXI: Modernização No Cultivo e Diversificação Do Uso Da Erva-Mate*, 34–37.
- Wendling, I., Dutra, L. F., & Grossi, F. (2007). Produção e sobrevivência de miniestacas e minicepas de erva-mate cultivadas em sistema semi-hidropônico. *Pesquisa Agropecuária Brasileira*, 42(2), 289–292. <https://doi.org/10.1590/s0100-204x2007000200019>
- Wendling, I., Sturion, J. A., Stuepp, C. A., Reis, C. A. F., Ramalho, M. A. P., & de Resende, M. D. V. (2018). Early selection and classification of yerba mate progenies. *Pesquisa Agropecuaria Brasileira*, 53(3), 279–286. <https://doi.org/10.1590/S0100-204X2018000300002>
- Westphalen, D. J., Angelo, A. C., Rossa, Ü. B., Helm, C. V., Radetski, C. M., & Gomes, E. N. (2022). *Phytochemical composition of yerba mate leaves (Ilex paraguariensis) and its relation with cultivation conditions*. 2020, 99–107.
- Workman, J. . J., & Weyer, L. (2007). Practical Guide to Interpretive Near-Infrared Spectroscopy. In *Practical Guide to Interpretive Near-Infrared Spectroscopy*. <https://doi.org/10.1201/9781420018318>
- Zwanenburg, G., Hoefsloot, H. C. J., Westerhuis, J. A., Jansen, J. J., & Smilde, A. K. (2011). ANOVA–principal component analysis and ANOVA–simultaneous component analysis: A comparison. *Journal of Chemometrics*, 25(10). <https://doi.org/10.1002/cem.1400>

Material Suplementar

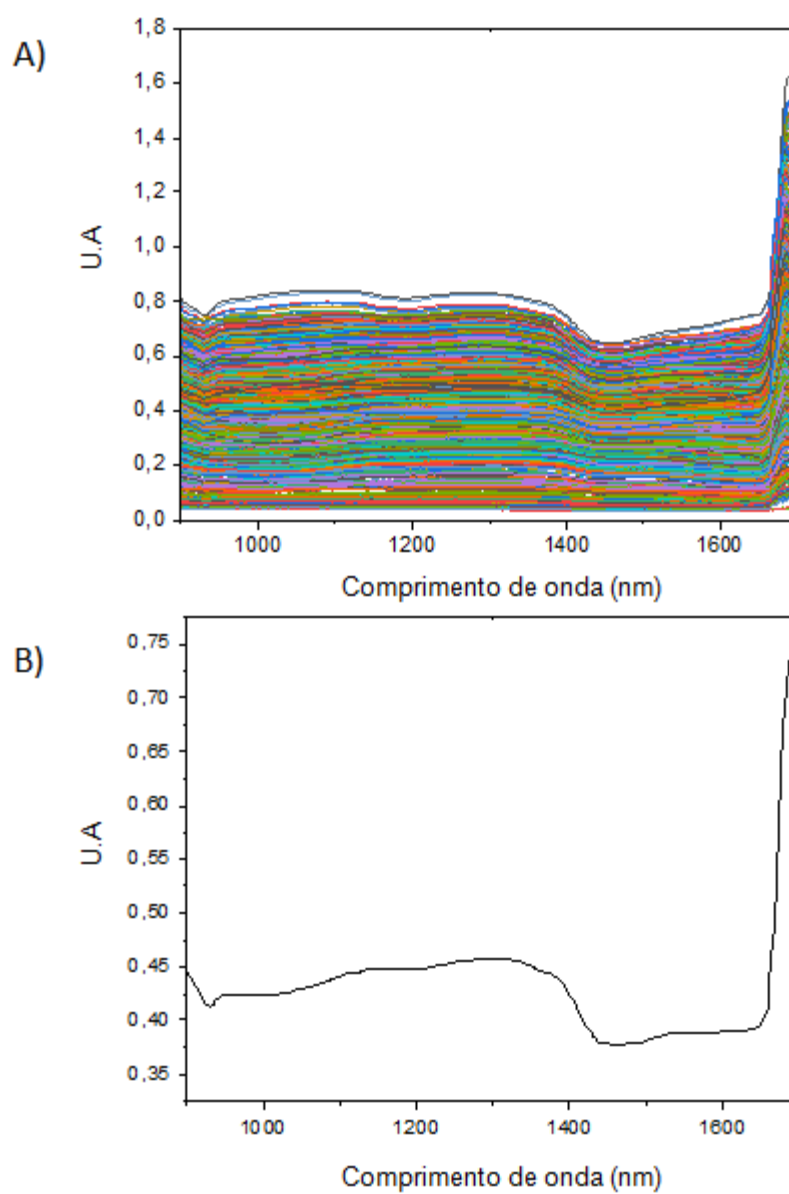


Fig. S1. Mil espectros NIR adquiridos pela leitura direta no material vegetal de amostras descafeinadas (a) seguidos pela média (b).

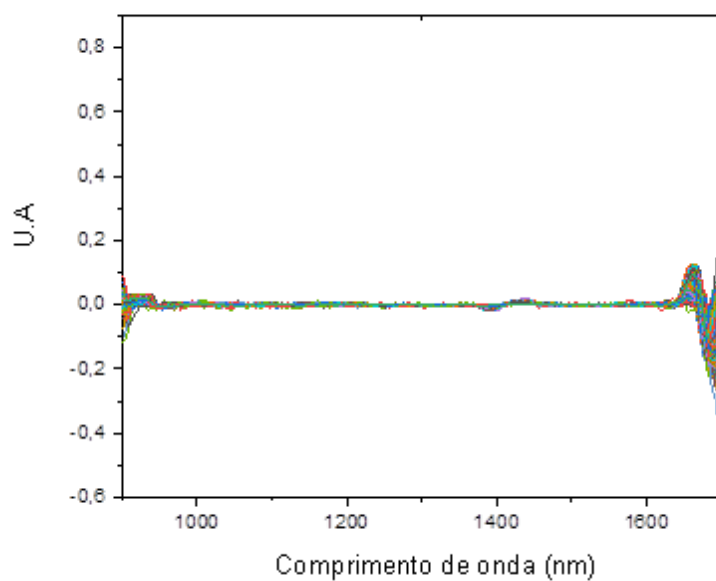


Fig. S2. Mil espectros NIR pré-tratados com SNV seguido pela segunda derivada de Savitzky-Golay (janela de 11 pontos e polinômio interpolador de segundo grau).

Manuscrito publicado em Food Analytical Methods, v. 16, p. 469–477, 2023

Uso da Espectroscopia NIR portátil para identificação de clones de Ervamate (*Ilex Paraguariensis*) quimicamente diferenciados

Andressa Gomes de Almeida^a, Elis Daiane Pauli^a, Cláudia Domiciano Tormena^a, Ivar Wendling^b, Miroslava Rakocevic^b, Roy Edward Bruns^c, Ieda Spacino Scarminio^{,a}, Gustavo Galo Marcheafave^{*,c}*

^a *Laboratório de Quimiometria em Ciências Naturais (LQCN), Departamento de Química, Universidade Estadual de Londrina, P.O.Box 6001, 86051-990 Londrina-PR, Brasil.*

^b *Embrapa Florestas, P.O.Box 319, 83411-000 Colombo-PR, Brasil.*

^c *Instituto de Química, Universidade Estadual de Campinas, P.O.Box 6154, 13083-970 Campinas-SP, Brasil.*

*ieda@uel.br (I.S. Scarminio), ggalo@unicamp.br (G.G. Marcheafave)

Resumo

As plantas de erva-mate fazem parte da floresta ombrófila mista com araucárias. Dois grupos de clones de erva-mate geneticamente melhorados, produzidos de acordo com o interesse de um público consumidor seletivo, foram classificados como: plantas com alto teor de cafeína e descafeinadas. Estratégias para o controle de qualidade deste tipo de insumo alimentício são essenciais e ferramentas multivariadas podem auxiliar neste procedimento. O tratamento por meio da Análise Discriminante de Mínimos Quadrados Parciais (do inglês, *Partial least squares-discriminant analysis*, PLS-DA) em dados de refletância (900–1700 nm), obtidos com a espectroscopia de infravermelho próximo (do inglês, *Near Infrared Spectroscopy*, NIRS), foi usado para identificar plantas de erva-mate quimicamente diferenciadas, com análise direta no material vegetal. As plantas de erva-mate foram cultivadas no sistema semi-hidropônico sob cinco níveis de sombreamento (0%, 40%, 51%, 76% e 82%). A robustez do modelo matemático foi verificada para todas as plantas nos diferentes níveis de sombreamento. O modelo PLS-DA apresentou uma sensibilidade de 96,52% para o conjunto de treinamento e 93,33% para o conjunto de teste. Especificidade superior a 97,12% foi encontrada para ambos os conjuntos, com taxa de eficiência de 96,82% para o conjunto de treinamento e 95,31% para o conjunto de teste. O teste de Wilcoxon e os testes de randomização mostraram um excelente ajuste do modelo. A análise de componentes principais (do inglês, *Principal Component Analysis*, PCA) dos espectros NIR demonstrou que o sombreamento afetou mais a composição química nos clones com alto teor de cafeína do que nos descafeinados. Isso indica que a síntese metabólica de cafeína em plantas de erva-mate representa uma estratégia adaptativa a condições de luz elevada.

Palavras-chave: Erva-mate geneticamente melhorada, Cafeína, Condições de luz, NIRS, PLS-DA, PCA.

1. Introdução

A erva-mate (*Ilex paraguariensis* St. Hil., Aquifoliaceae) faz parte da floresta ombrófila mista com araucárias e se encontra amplamente associada a espécie *Araucaria angustifolia* (Bert.) O. Kuntze. As folhas e galhos desta planta possuem ampla aplicação comercial que vem sendo explorada desde as primeiras décadas do século XIX (Brandt & Silva, 2014). Atualmente a produção da erva-mate se dá sob os sistemas de monocultura e agroflorestal, acarretando em variações morfológicas (Rakocevic & Martim, 2011), fisiológicas (Rakocevic et al., 2011; Takeshi Matsunaga et al., 2014) e metabólicas (Tormena et al., 2020) decorrentes da incidência luminosa.

O Brasil constitui-se como o maior produtor mundial de erva-mate, possuindo cerca de 450.000 km² de território onde há ocorrência natural da erva (Bonfatti Júnior et al., 2018) e além do Brasil, países como a Argentina e o Paraguai também possuem erva-mate nativa, porém, em menor extensão (Wrege et al., 2020). Essa planta, que é mundialmente conhecida devido ao seu consumo sob as formas de chimarrão, tererê e chá-mate, vem ganhando espaço em variadas áreas da indústria e pesquisa em função da sua composição química rica em fitoquímicos como: metilxantinas, polifenóis e saponinas. Estes biocompostos possuem atividades estimulantes, diuréticas, estão associados à prevenção de cânceres, danos no DNA e doenças cardiovasculares (Mazur et al., 2014) e alguns estudos exploram seu potencial como antioxidante, antimicrobiano (Oh et al., 2013) e antibacteriano (Fayad et al., 2020).

Com relação aos metabólitos secundários provenientes da erva-mate (*I. paraguariensis*), estudos indicam majoritariamente a presença de teobromina, teofilina e cafeína, assim como ácidos clorogênicos e dois de seus isômeros: ácido 3-cafeoilquínico e 4-cafeoilquínico (Negrin et al., 2019). A cafeína induz efeitos estimulantes, que reduzem a fadiga, aumentam o estado de alerta e diminuem o tempo de reação. Estudos ainda apontam que a ingestão em doses moderadas pode reduzir o risco de fibrose hepática, cirrose, câncer, depressão e doença de Parkinson, além de melhorar sutilmente a função pulmonar de adultos e auxiliar no tratamento de apneia em bebês prematuros. Entretanto, pode levar a um quadro de ansiedade (em indivíduos com pré-disposição), insônia, nervosismo, disforia, tremores, aumentar a pressão sanguínea, prejudicar o crescimento de fetos ou

umentar o risco de aborto (van Dam et al., 2020). Tendo em vista os diversos fatores, tanto positivos quanto negativos, a produção de alimentos/bebidas com diferentes dosagens de cafeína se torna uma proposta atraente para um mercado que está constantemente mudando e buscando atender as variadas necessidades dos consumidores. Algumas técnicas convencionais para remoção de cafeína já vêm sendo empregadas, tal como: a descafeinação por água, extração por solvente e por dióxido de carbono supercrítico (Saldaña et al., 2000), entretanto a quantidade de resíduos poluentes gerados torna a prática preocupante. Outras técnicas como: degradação enzimática e por microorganismos estão sendo desenvolvidas, porém ainda sem muito conhecimento a respeito do possível aumento da concentração de substâncias tóxicas decorrente de mudanças nas rotas biossintéticas (Gokulakrishnan et al., 2005).

O interesse na aplicação e desenvolvimento de técnicas para melhoramento genético em organismos como plantas vêm crescendo paralelo ao avanço da tecnologia, fomentando pesquisas e potencializando plantações de uso agrícola. Essas modificações podem ocorrer por DNA recombinante, edição de genoma, metilação de DNA dirigida por RNA, cisgênese e intragênese, transenxertos, agroinfiltração, indução haplóide e reprodução acelerada (Eckerstorfer et al., 2019). Na erva-mate, a partir da década de noventa, têm sido empregadas técnicas de modificação genética com o objetivo de adaptação, produção de massa verde, resistência a pragas e doenças, desfolhamento, tipo de ramificação/ arquitetura (Resende et al., 2000). Alguns estudos têm sido realizados com a finalidade de avaliar os efeitos físicos, químicos e biológicos causados pelas condições de plantio, tal como o nível de exposição à radiação solar (Caron et al., 2014; Rakocevic et al., 2011; Rossa et al., 2017). O melhoramento genético permitiu a produção de plantações com níveis variados de cafeína (Wendling, 2016). Sendo uma abordagem recente, não foram encontrados estudos relacionados a regulamentação/validação de erva-mate modificada geneticamente, bem como métodos para avaliação da qualidade desses novos produtos.

Quanto a comercialização da erva-mate processada, já foram encontradas adulterações decorrentes da mistura de folhas pertencentes a outras espécies de *Ilex*, como: *I. dumosa*, *I. pseudobuxus*, *I. brevicuspis*, *I. theezans* e *I. microdonta* (Giberti, 1989). Esta adulteração promove modificação da composição química e, conseqüentemente, de importantes propriedades sensoriais da bebida, tal como o

sabor (Pires et al., 1997). Ainda foi observado que a adulteração com folhas de *Ilex dumosa* aumenta a concentração de saponinas hemolíticas, que afetam severamente as células vermelhas presentes no sangue (Zohra & Fawzia, 2014), enquanto que a presença de *P. sellowii* (pessegueiro bravo) atribui heterosídeos cianogênicos à composição, que apresentam alta toxicidade relatada (Valduga et al., 1997).

O desenvolvimento de métodos para o controle de qualidade de insumos alimentícios é imprescindível e, sendo assim, a busca por técnicas capazes de fornecer resultados rápidos e confiáveis, tal como métodos instrumentais de análise direta, tem se intensificado, tendo em vista resultados promissores em pesquisas recentes (Silva et al., 2021; Tormena et al., 2020). A Espectroscopia no Infravermelho Próximo (do inglês, Near infrared, NIR) tem sido empregada por se tratar de uma técnica não destrutiva, além de fornecer resultados rápidos sem a necessidade de exaustivas manipulações de amostra (Pasquini, 2003). Aliado a tratamentos quimiométricos adequados, como as ferramentas PLS-DA (do inglês, Partial least squares – Discriminant Analysis), PCA (do inglês, Principal Component Analysis), ASCA (do inglês, ANOVA-simultaneous component analysis) é possível extrair a maior quantidade possível de informações presentes nos dados (Marcheafave, Pauli, et al., 2020; Tormena et al., 2019).

Este trabalho objetiva avaliar a potencial discriminação/classificação de folhas de erva-mate descafeinadas e com alto teor de cafeína obtidas a partir de um programa de melhoramento genético, por meio da análise direta do material vegetal com um instrumento portátil NIR e modelos PLS-DA. Adicionalmente, comparar o efeito de vários níveis de sombreamento na composição química de folhas com ausência e presença de cafeína pelos espectros NIR e PCA.

2. Materiais e Métodos

2.1 Produção da planta e amostragem

Foram utilizadas folhas de erva-mate pertencentes a clones com alto teor de cafeína (genótipos 4–8-3 com 2,35 g 100 g⁻¹ de cafeína) e descafeinado (genótipos 7–61-3 com 0,02 g 100 g⁻¹ de cafeína) produzidos a partir do programa de melhoramento genético da Embrapa Florestas (Sturion et al., 2017; Wendling et al., 2018). A categorização da cafeína foi realizada de acordo com Schuhli e outros

(Schuhli et al., 2019), onde o índice das amostras com alto teor de cafeína é superior a 1,6% (16 mg g⁻¹) e o das amostras descafeinadas menor que 0,1% (1 mg g⁻¹). O teor de cafeína foi determinado com extração em água quente de acordo com Helm et al. (Vieira Helm et al., 2015).

As plantas foram cultivadas em estufa (temperaturas mínimas e máximas de 13 °C e 37 °C, respectivamente, com média de 25,3 °C e ≥ 85% de umidade relativa do ar) na cidade de Colombo, Estado do Paraná, Brasil (25°20'S e 49°14'W, 950 m de altitude). Eles foram cultivados em um espaço adensado com 44 plantas por metro quadrado (espaçamento de 15 cm × 15 cm) em um sistema de canal semi-hidropônico com leitos de areia (Wendling et al., 2007) sob cinco níveis de sombreamento (0%, 40%, 51%, 76% e 82%), obtido usando telas comerciais de polietileno (sombras) sobre os canais. Antes da coleta de folhas, as plantas sofreram 11 amostragens ao longo de 2 anos para outros tipos de análises. As folhas coletadas para cada clone e nível de sombreamento foram secas em forno de micro-ondas (Tomasi et al., 2021) e armazenados para estudos posteriores.

2.2 *Medições NIR*

A aquisição dos dados no modo de refletância NIR foi realizada com medições espectrais diretas nas folhas usando o Software DLP NIRscan Nano (Texas Instruments Inc., Dallas, TX, EUA) contra um branco a 25 ± 1 °C. Os espectros de reflectância referentes ao padrão de cafeína (SIGMA-ALDRICH) foram obtidos comprimindo o sólido entre duas lâminas de microscópio (26 × 76 – 0,9 a 1,1 mm). Cada espectro foi obtido pela média de 99 varreduras entre 900 e 1700 nm com uma resolução de 3,5 nm. Para cada folha, dez espectros foram adquiridos em diferentes posições de uma lâmina foliar.

2.3 *Tratamentos Quimiométricos*

2.3.1 *Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA)*

A PLS-DA é um método supervisionado para discriminar dados multivariados. O classificador recebe valores numéricos de entrada para cada amostra, como 0 e 1, e a matriz **X** é calibrada contra este valor de entrada numérica. Consequentemente, um valor estimado é obtido, e o modelo determina a classe de adesão para cada

amostra (limiar/limite de decisão) (Brereton, 2016). Finalmente, o modelo é avaliado a partir de uma série de figuras de mérito para validar a robustez da discriminação (Nunes et al., 2019). O conjunto de treinamento foi dimensionado em 1400×228 , de modo que a matriz de dados continha 1400 espectros de refletância NIR igualmente distribuídos entre as duas classes de folhas de erva-mate e 228 valores de refletância NIR. A dimensão da matriz do conjunto de teste foi 600×228 contendo 300 espectros de cada classe.

2.3.2 *Análise de componentes principais (PCA)*

A PCA é uma análise exploratória de dados, sendo o método quimiométrico mais aplicado em dados químicos multivariados (Bro & Smilde, 2014). Em geral, a PCA envolve uma transformação matemática da matriz de dados químicos (\mathbf{X} ; matriz formada por valores de absorbância, refletância, entre outros) em novas matrizes chamadas scores (\mathbf{T}) e loadings (\mathbf{P}), e também o erro de modelagem (\mathbf{E}) (Brereton, 2016). É possível calcular quantas matrizes de scores e loadings quisermos, desde que a dimensão não exceda a menor dimensão da matriz original (\mathbf{X}). Em química, geralmente o número de dimensões é determinada a partir da interpretação das informações extraídas das matrizes \mathbf{P} e \mathbf{T} . As dimensões da matriz utilizada na PCA deste trabalho foram de 1000×228 para cada uma das duas classes de folhas de erva-mate (clones com alto teor de cafeína e descafeinados). O número de linhas (1000) representou diferentes folhas coletadas, enquanto as colunas (228) representavam os valores de refletância NIR. A PCA foi aplicada para comparar o efeito de diferentes níveis de sombreamento (cinco níveis) para cada classe de amostra e comparar graficamente a dispersão dos escores.

2.4 *Software e Análises estatísticas*

Todos os cálculos foram realizados no software Matlab 2016® (R2016b) usando ferramentas da caixa PLS8.7.1® da Eigenvector Research, enquanto o OriginPro 2018 foi aplicado para plotar figuras de loadings.

3. Resultados e Discussão

Inicialmente, procedimentos de pré-processamento e detecção de outliers

foram realizados no conjunto de dados NIR com foco na qualidade da construção do modelo PLS-DA. Varias técnicas de pré-processamento foram testadas com a melhor sendo a primeira derivada com Savitzky-Golay (polinômio de primeira ordem com uma janela de 11 pontos), e por fim, os dados foram centrados na média. O perfil metabólico pré-tratado obtido entre 900 e 1700 nm é mostrado na Fig. 1. Nesta região do infravermelho, sobreposição e combinação de bandas de vários compostos de erva-mate são encontrados com bandas características correspondentes a grupos metil (CH_3) e grupo C-H em 1450 e 1130 nm, um sinal geralmente associado com a presença de lipídios (em torno de 1660 nm) entre outros (Burns & Ciurczak, 2009). Por meio da média dos espectros NIR pré-processados obtidos pra as folhas de erva-mate com alto teor de cafeína e descafeinadas é possível verificar visualmente que as principais diferenças espectrais ocorrem entre 1650 e 1700 nm (Fig. S1).

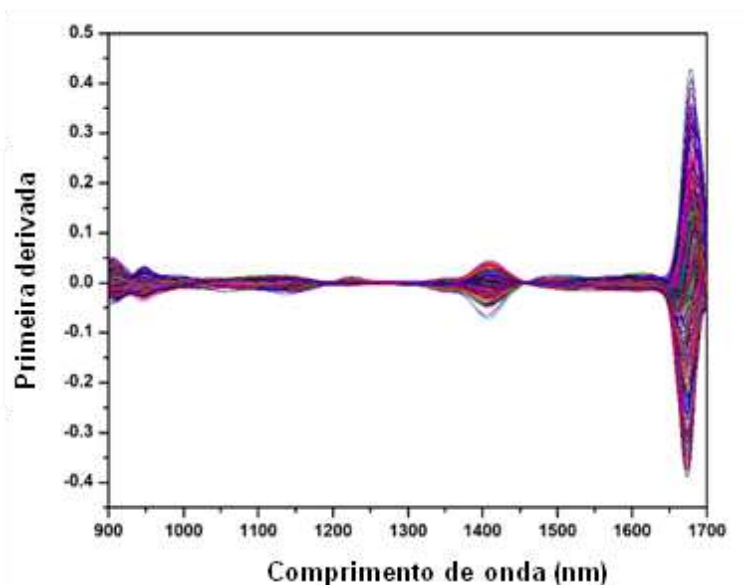


Fig. 1. Espectro NIR de folhas de erva-mate após pré-tratamento com primeira derivada de Savitzky-Golay centrado na média.

Para o desenvolvimento de duas classes do modelo PLS-DA, variáveis de referência fictícias foram definidas no vetor y como “1” para folhas de erva-mate descafeinada e “0” para folhas de erva-mate com alto teor de cafeína. As amostras foram divididas em conjunto de treinamento (70%) e conjunto de teste (30%), usando o algoritmo de Kennard e Stone para cada classe (Kennard & Stone, 1969). A matriz de dados continha 700 espectros de treinamento e 300 espectros de teste para folhas de erva-mate descafeinadas, e 700 espectros de treinamento e 300

espectros de teste para folhas com alta teor de cafeína. Para o desenvolvimento do modelo PLS-DA, uma análise para identificação de amostras anômalas (outliers) foi realizada, relacionando T^2 de Hotelling *versus* Q residual. T^2 de Hotelling está relacionado com medidas de distância entre a amostra e o centro dos dados, enquanto Q residual representa os resíduos não modelados nos espectros NIR. Neste caso, 81 amostras originais do conjunto de calibração, representando 4,05%, apresentaram valores de T^2 de Hotelling e Q residuais respectivamente elevados. Então, com esse comportamento anômalo, as respectivas amostras foram retiradas, garantindo a homogeneidade dos dados no espaço multidimensional (Fig. 2).

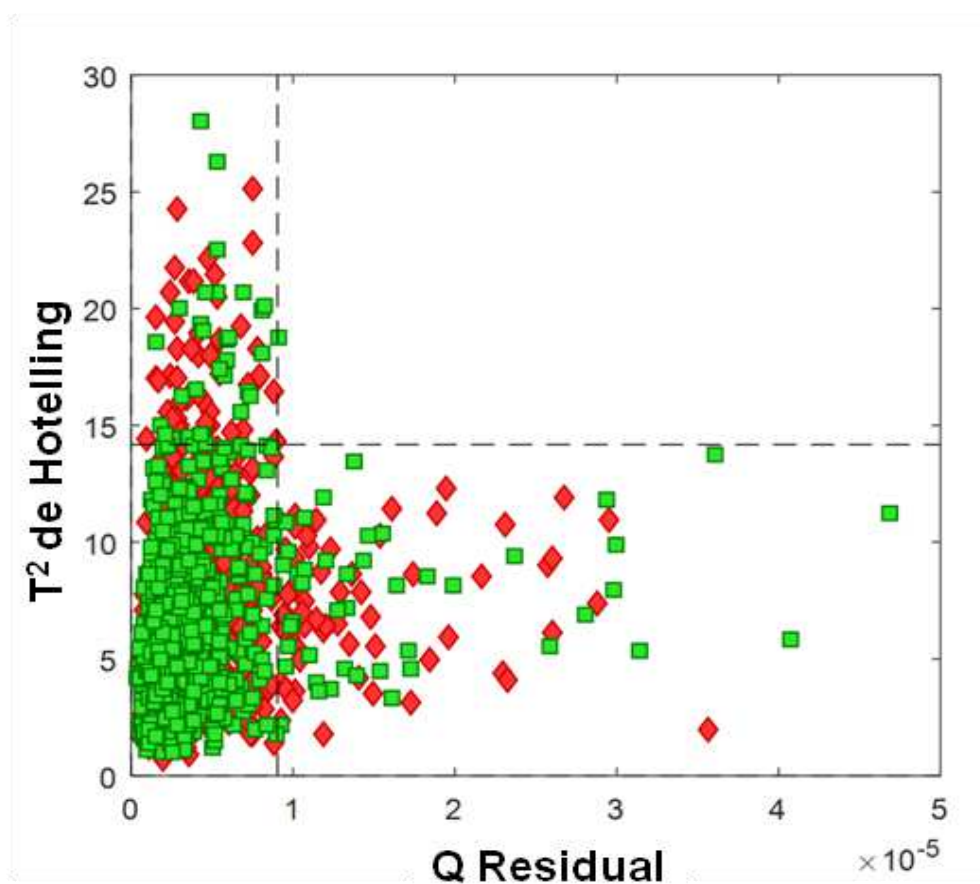


Fig. 2. Q residual contra T^2 de Hotelling para o modelo de discriminação PLS-DA entre plantas de erva-mate quimicamente diferenciadas (clones com alto teor de cafeína (■) e descafeinados (◆)) sob cinco níveis de luz usando espectroscopia de infravermelho próximo na região de 900-1700 nm.

Por meio do erro de classificação de validação cruzada (do inglês, Cross-Validated Classification Error, CVCE), sete variáveis latentes foram escolhidas, a partir da validação cruzada do tipo venezianas com 10 divisões. O modelo PLS-DA

considerado mais adequado conteve 99,88% de variação do bloco X e 73,85% do bloco Y. Os valores de previsão Y para plantas de erva-mate descafeinadas (a) e plantas com alto teor de cafeína (b), ambas sob cinco níveis de sombreamento, podem ser observados na Fig. 3. O limite bayesiano de 0,50 estimado pelo modelo é representado pela linha vermelha pontilhada. Os gráficos previstos de Y ilustram a distribuição de amostras entre as classes como previsto pelo modelo e apontam discriminações incorretas, com falsos positivos e falsos negativos (Fig. 3). Para o conjunto de treinamento composto por plantas de erva-mate descafeinadas investigadas sob cinco níveis de sombreamento (Fig. 3a), 19 amostras de folhas descafeinadas foram classificadas como folhas com alto teor de cafeína (falso-negativo), enquanto 23 amostras foram classificadas como descafeinadas, embora suas folhas tenham alto teor de cafeína (falso-positivo). A partir destas classificações, a taxa de falsos positivos (FPR) correspondeu a 3,46% e a taxa de falsos negativos (FNR) a 2,89% para o conjunto de treinamento. Assim, a sensibilidade foi de 96,54% (Tabela 1). Para o conjunto de teste, as taxas de FPR e FNR corresponderam a 2,69% e 6,67%, respectivamente, e a sensibilidade foi de 93,33% para o conjunto de teste. A especificidade correspondeu a 97,11% e 97,31% para os conjuntos de treinamento e teste, respectivamente. Finalmente, a taxa de eficiência foi de 96,82% para o conjunto de treinamento e 95,32% para o conjunto de teste.

O coeficiente de correlação de Matthews foi de 0,936, demonstrando a alta qualidade da previsão (Chicco & Jurman, 2020). O teste-t de randomização no nível de 95% indicou que os modelos permutados e não permutados foram significativamente diferentes. A classificação do modelo original PLS-DA foi diferente de uma classificação aleatória entre as folhas de erva-mate descafeinada e aquelas com alto teor de cafeína. O teste de Wilcoxon (teste de classificação sinalizado de Wilcoxon), teste de sinais (teste de classificação de sinais em pares) e Rand t-test (teste-t de randomização) mostrou que o modelo é não sobreajustado com valores menores que 0,05.

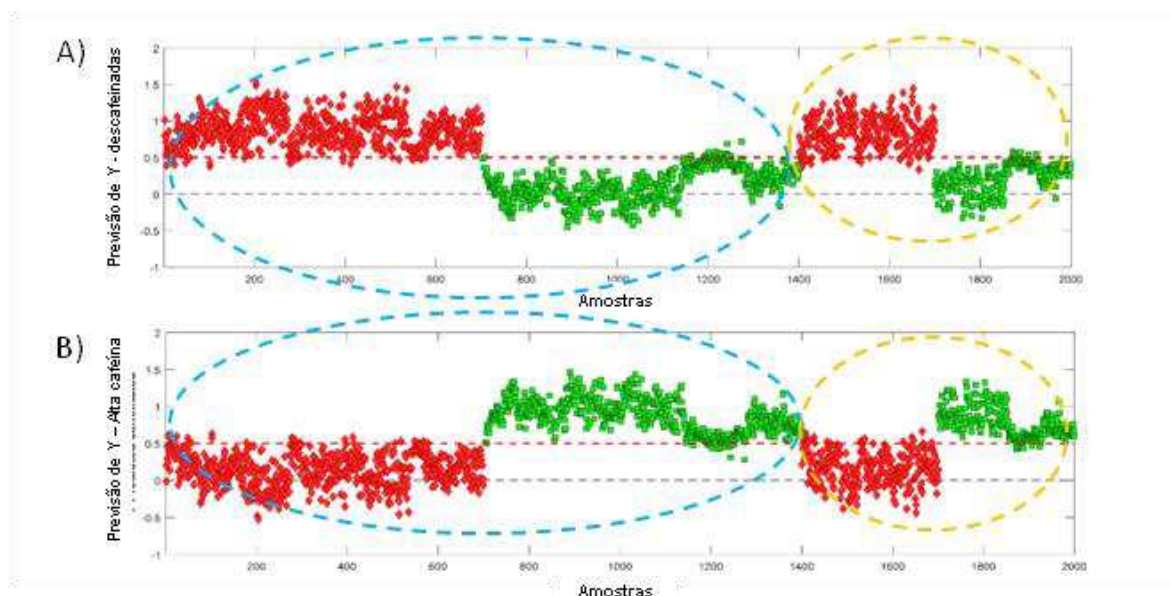


Fig. 3. Modelo de previsão PLS-DA para clones de erva-mate descafeinados (♦) (a) e clones com alto teor de cafeína (■) (b) sob cinco níveis de luz, obtidos pelo uso da espectroscopia NIR portátil na região de 900-1700 nm. Modelo PLS-DA construído com 1400 amostras de treinamento (primeiro grupo) e 600 amostras para validação (segundo grupo).

Tabela 1. Figuras de mérito para o modelo PLS-DA de folhas de erva-mate.

Figuras de mérito	Conjunto de Treinamento	Conjunto de teste
Sensibilidade	96.52%	93.33%
Especificidade	97.12%	97.31%
Taxa de falso-positivo	3.46%	2.69%
Taxa de falso-negativo	2.89%	6.67%
Taxa de eficiência	96.82%	95.31%

Para interpretar a classificação espectral das amostras no modelo PLS-DA, os scores da importância da variável na projeção (do inglês, Variable Importance in Projection, VIP) mostraram a importância estimada de cada variável na projeção da classificação (Nunes et al., 2019). Variáveis com VIP maior que 1 são consideradas significativas/necessárias na construção do modelo. Os comprimentos de onda essenciais neste modelo foram 1395, 1660, 1682 e 1690 nm (Fig. 4). A banda de

1395 nm caracteriza a banda de vibração C-H da estrutura de $-CH_2$ (Burns & Ciurczak, 2009). As bandas em 1660 e 1690 nm caracterizam a ligação C-H do primeiro sobretom para as estruturas de $-CH_3$ (Burns & Ciurczak, 2009). A banda em 1682 nm caracteriza o estiramento da ligação C-H do primeiro sobretom para estruturas aromáticas (Burns & Ciurczak, 2009). Essas bandas na região NIR são características da cafeína (Fig. S2).

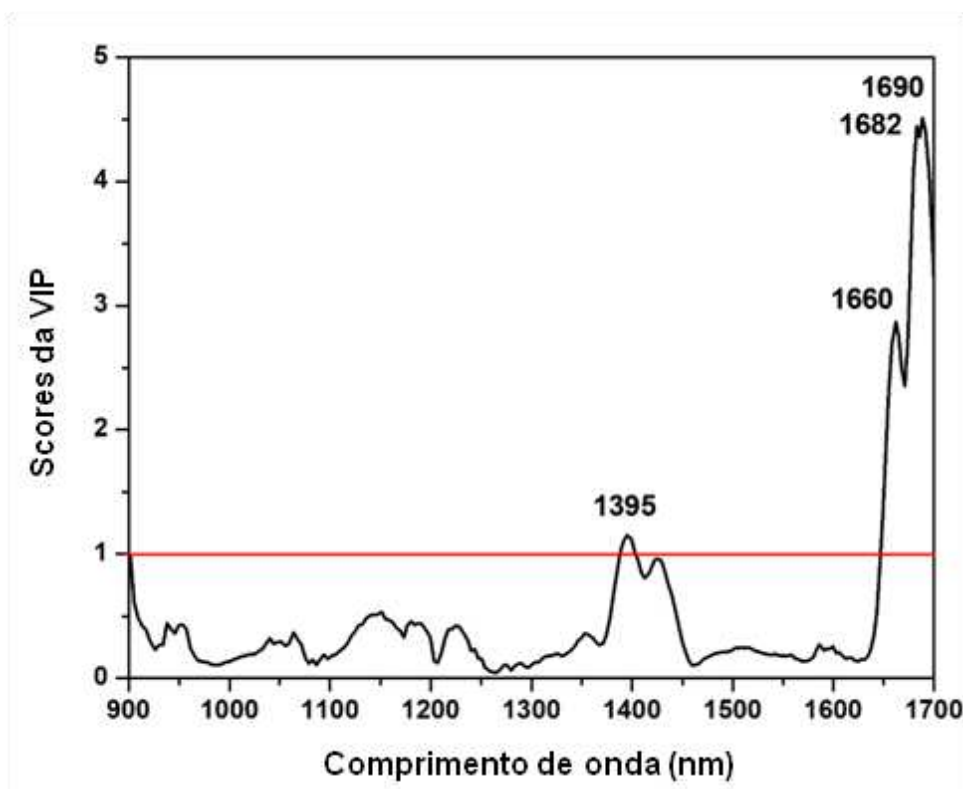


Fig. 4. Scores da importância da variável na projeção (VIP) no modelo discriminatório PLS-DA para plantas de erva-mate com alto teor de cafeína e descafeinadas sob cinco níveis de luz resultantes do uso da espectroscopia NIR portátil na região de 900-1700 nm.

Diante dos resultados, pode-se concluir que os diferentes níveis de luz não causaram desvios significativos no modelo de classificação PLS-DA. Desta forma, a robustez do modelo com relação a diferentes incidências luminosas foi confirmada e este pode ser aplicado, por exemplo, para diferentes níveis de luz do perfil vertical da planta. Mudanças químicas devidas à estratificação já foram encontrados em grãos de *Coffea arabica* e indicam mudanças químicas decorrentes de diferentes intensidades de irradiação solar incidente (Marcheafave et al., 2021; Rakocevic et al., 2018). Esse tipo de modificação ambiental, se não estiver incluído nos modelos

de classificação, pode distorcer os resultados quando aplicados em diferentes condições edafoclimáticas (Marcheafave et al., 2021), já que a classificação é baseada em um perfil metabólico que abrange o metabolismo secundário das plantas, sensível às variações ambientais.

Para investigar a variação dos perfis químicos a partir de NIR das amostras obtidas sob diferentes condições de luz, a matriz espectral foi submetida a PCA, separadamente para cada classe de clones de cafeína (Fig. 5). Para avaliar a separação dos dados, uma matriz contendo a PCA vs as amostras foi gerada (Fig S1)

Os scores da PC4 para os espectros das folhas de erva-mate descafeinadas foram parcialmente separadas com 0,72% da variância dos dados (Fig. 5A). Na região positiva estavam os terceiro e quarto níveis de luz, enquanto na região negativa situaram-se os perfis dos primeiro e segundo níveis de luz. Na região positiva é possível observar loadings em 1343 nm (combinação de CH na estrutura do CH₃), 1405 nm (O-H primeiro sobretom na estrutura ROH), e 1660 nm (C-H do primeiro sobretom para estruturas -CH₃) (Fig. 6A) (Burns & Ciurczak, 2009). Vale ressaltar que, além da componente PC4 explicar menos de 1% da variação total, os níveis menores e maiores de incidência luminosa apresentam-se extensivamente sobrepostos. Assim, a Fig. 5A mostra quase nenhuma separação para plantas descafeinadas sob diferentes condições de luz quando analisados nesta região espectral. A PCA para os espectros NIR dos clones de erva-mate com alto teor de cafeína (Fig. 5B) mostrou separação de escores melhor definida quando comparado com os clones descafeinados (Fig. 5A).

A PC2, com 8,75% de variância dos dados, apresentou os scores para o quarto e quinto níveis de luz na região positiva e os scores para o primeiro, segundo e terceiro níveis de luz na região negativa (Fig. 5B). Por meio dos gráficos de loadings relacionados aos clones de alto teor de cafeína (Fig. 6B), a separação da região positiva ocorreu pelas bandas em 1405 (O-H primeiro sobretom na estrutura ROH) e 1686 nm (estiramento C-H do primeiro sobretom para estruturas aromáticas) (Burns & Ciurczak, 2009). Na região negativa, a banda em 1660 nm (C-H do primeiro sobretom para estruturas -CH₃) foi a responsável para a separação (Burns & Ciurczak, 2009). Provavelmente a maior separação que ocorreu devido aos diferentes níveis de luz para as folhas de erva-mate com alto teor de cafeína foi devido ao metabolismo secundário deste alcalóide (Fig. S2), uma vez que existe

semelhança entre os valores de loadings com os obtidos no padrão analítico de cafeína (Fig. S3). Embora a literatura apresente resultados contraditórios com relação a concentração de cafeína em plantas de erva-mate submetidas a diferentes níveis de luz (Rakocevic et al., 2012; Riachi et al., 2018), ficou evidente que este metabólito está relacionado ao mecanismo de defesa e aclimação da planta em resposta a condições de estresse.

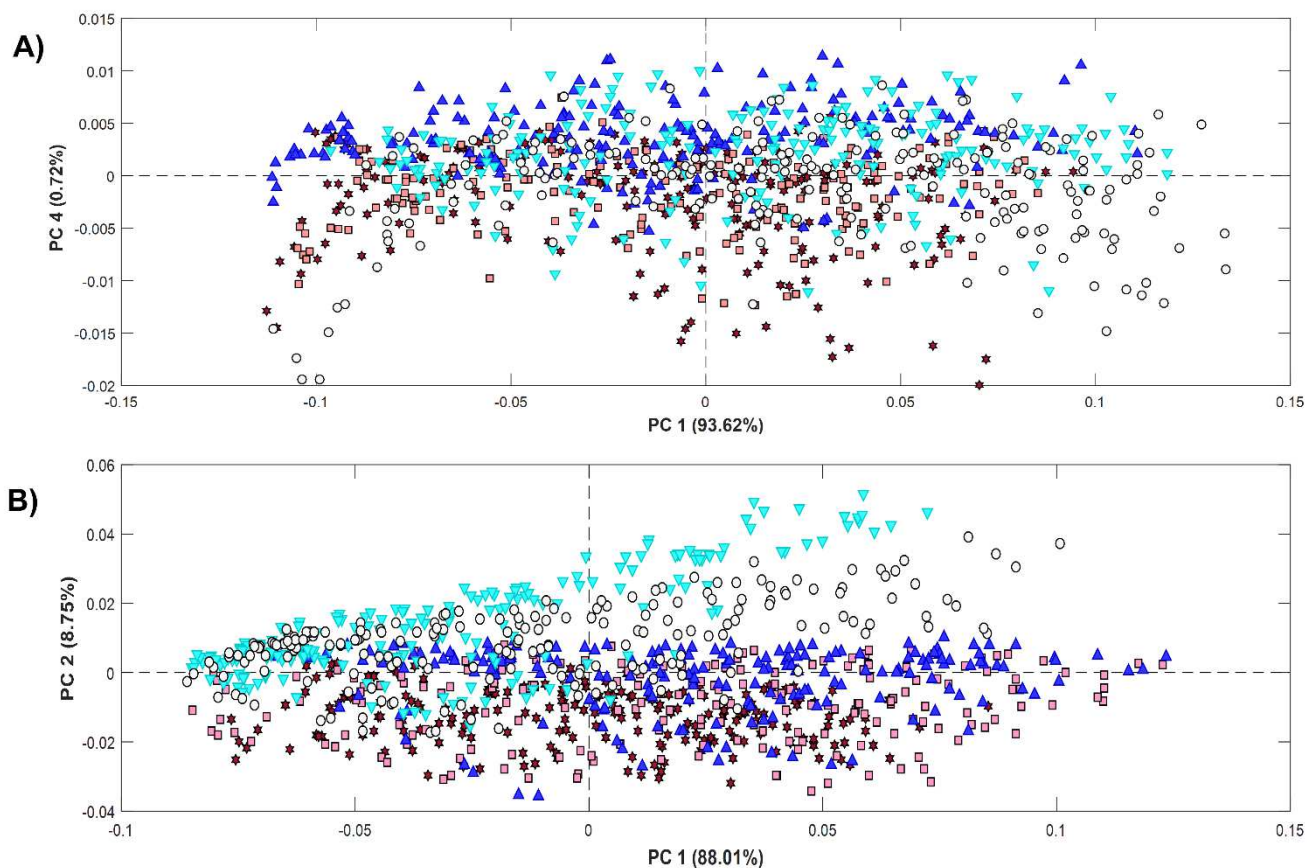


Fig. 5. Gráficos de scores resultantes da análise de componentes principais para folhas de erva-mate descafeinadas (a) e com alto teor de cafeína (b) em plantas cultivadas sob cinco diferentes níveis de luz (0%, 40%, 51%, 76%, e 82%), obtidos a partir de 1000 medições NIR na faixa de 900-1700 nm para cada classe. (★) primeiro nível (0%), (■) segundo nível (40%), (▲) terceiro nível (51%), (▼) quarto nível (76%), e (○) quinto nível de sombreamento (82%).

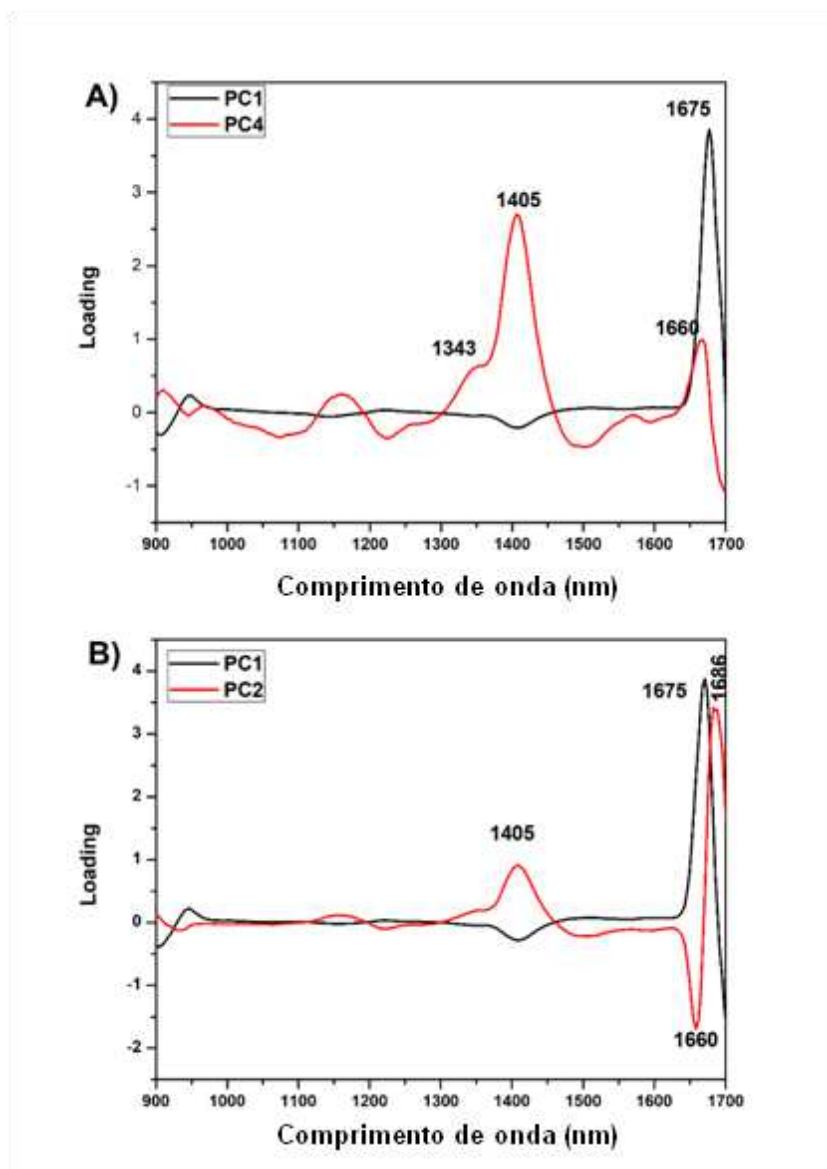


Fig. 6. Loadings *versus* comprimento de onda na região NIR para análise exploratória de folhas de erva-mate em clones descafeinados (a) e com alto teor de cafeína (b) cultivados sob cinco diferentes níveis de luz.

Conclusão

Este estudo indica que o modelo de classificação PLS-DA associado a análise NIR não destrutiva forneceu uma ferramenta efetiva para o controle de qualidade de plantas de erva-mate quimicamente diferenciadas. Todas as figuras de mérito avaliadas para a classificação da erva-mate descafeinada e com alto teor de cafeína demonstraram a construção de um excelente modelo com sensibilidade acima de 93%, especificidade acima de 97% e taxa de eficiência acima de 95% para os conjuntos de calibração e teste. Tanto o teste de Wilcoxon quanto o teste-t

randomizado mostraram que o modelo não é sobreajustado. Os diferentes níveis de sombreamento não afetaram o modelo de classificação. Assim, o modelo aqui demonstrado tem alta confiabilidade na identificação de folhas de erva-mate quimicamente diferenciadas independente da posição vertical na planta. A análise exploratória dos espectros NIR por PCA mostrou que a variação metabólica nas folhas entre os diferentes níveis de sombreamento foram muito mais significativos para os clones com alto teor de cafeína do que os descafeinados, sugerindo a importância do papel da cafeína na aclimatação para elevados níveis de luz.

Agradecimentos

ISS (Ieda Spacino Scarminio) e MR (Miroslava Rakocevic) agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelas bolsas concedidas (PQ, Processo: 302204/2018-0; PV, Processo: 350509/2020-4). GGM (Gustavo Galo Marcheafave) agradece a bolsa 2020/11463-5, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). A EDP (Elis Daiane Pauli) agradece a oportunidade de pós-doutoramento na Universidade Estadual de Londrina.

Referências

- Bonfatti Júnior, E. A., Lengowski, E. C., & Ludka Júnior, A. (2018). Mapeamento Do Processo Produtivo De Erva-Mate. *Revista Internacional de Ciências*, 8(1), 82–98. <https://doi.org/10.12957/ric.2018.32500>
- Brandt, M., & Silva, N. S. (2014). A coleta da erva-mate pela população cabocla do vale do rio do peixe e oeste de santa catarina: apropriação privada da terra e rupturas (décadas de 1900 a 1940). *Sociedade & Natureza*, 26(3), 459–469. <https://doi.org/10.1590/1982-451320140305>
- Brereton, R. G. (2016). Chemometrics : Data driven extraction for science: Second edition. In *Chemometrics: Data Driven Extraction For Science: Second Edition*. <https://doi.org/10.1002/9781118904695>
- Burns, D. A., & Ciurczak, E. W. (2009). Handbook of near-infrared analysis, 3rd ed. In *Analytical and Bioanalytical Chemistry* (Vol. 393, Issue 5).
- Caron, B. O., Schmidt, D., Manfron, P. A., Behling, A., Eloy, E., & Busanello, C. (2014). EFFICIENCY OF THE USE OF SOLAR RADIATION FOR PLANTS *Ilex paraguariensis*. *Ciência Florestal*, 24(2), 257–265.
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1). <https://doi.org/10.1186/s12864-019-6413-7>
- Eckerstorfer, M. F., Engelhard, M., Heissenberger, A., Simon, S., & Teichmann, H. (2019). Plants developed by new genetic modification techniques-Comparison of existing regulatory frameworks in the EU and Non-EU countries. *Frontiers in Bioengineering and Biotechnology*, 7(FEB). <https://doi.org/10.3389/fbioe.2019.00026>
- Fayad, E., El-Sawalhi, S., Azizi, L., Beyrouthy, M., & Abdel-Massih, R. M. (2020). Yerba Mate (*Ilex paraguariensis*) a potential food antibacterial agent and combination assays with different classes of antibiotics. *Lwt*, 125(March), 109267. <https://doi.org/10.1016/j.lwt.2020.109267>
- Giberti, G. C. (1989). Los parientes silvestres de la yerba mate y el problema de su adulteración. In *Dominguezia* (Vol. 7, Issue 1, pp. 1–21).
- Gokulakrishnan, S., Chandraraj, K., & Gummadi, S. N. (2005). Microbial and enzymatic methods for the removal of caffeine. *Enzyme and Microbial Technology*, 37(2), 225–232. <https://doi.org/10.1016/j.enzmictec.2005.03.004>
- Kennard, R. W., & Stone, L. A. (1969). Computer Aided Design of Experiments. *Technometrics*, 11(1). <https://doi.org/10.1080/00401706.1969.10490666>
- Marcheafave, G. G., Pauli, E. D., Tormena, C. D., Mattos, L. E., de Almeida, A. G., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2020). Irrigated and CO2 level effects on metabolism in *Coffea arabica* beans from mixture design – near infrared fingerprints. *Microchemical Journal*, 152(June 2019), 104276. <https://doi.org/10.1016/j.microc.2019.104276>
- Marcheafave, G. G., Tormena, C. D., Terrile, A. E., Salamanca-Neto, C. A. R., Sartori, E. R., Rakocevic, M., Bruns, R. E., Scarminio, I. S., & Pauli, E. D. (2021). 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. <https://doi.org/10.1016/j.foodchem.2021.129716>
- Mazur, L., Peralta-Zamora, P. G., Demczuk, B., & Ribani, R. H. (2014). Application of multivariate calibration and NIR spectroscopy for the quantification of methylxanthines in yerba mate (*Ilex paraguariensis*). *Journal of Food*

- Composition and Analysis*, 35(2), 55–60.
<https://doi.org/10.1016/j.jfca.2014.04.005>
- Negrin, A., Long, C., Motley, T. J., & Kennelly, E. J. (2019). LC-MS Metabolomics and Chemotaxonomy of Caffeine-Containing Holly (Ilex) Species and Related Taxa in the Aquifoliaceae. *Journal of Agricultural and Food Chemistry*, 67(19), 5687–5699. <https://doi.org/10.1021/acs.jafc.8b07168>
- Nunes, K. M., Andrade, M. V. O., Almeida, M. R., Fantini, C., & Sena, M. M. (2019). Raman spectroscopy and discriminant analysis applied to the detection of frauds in bovine meat by the addition of salts and carrageenan. *Microchemical Journal*, 147. <https://doi.org/10.1016/j.microc.2019.03.076>
- Oh, J., Jo, H., Cho, A. R., Kim, S. J., & Han, J. (2013). Antioxidant and antimicrobial activities of various leafy herbal teas. *Food Control*, 31(2), 403–409. <https://doi.org/10.1016/j.foodcont.2012.10.021>
- Pasquini, C. (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14(2), 198–219. <https://doi.org/10.1590/S0103-50532003000200006>
- Pires, V. S., Guillaume, D., Gosmann, G., & Schenkel, E. P. (1997). Saponins from Ilex dumosa, an Erva-maté (Ilex paraguariensis) Adulterating Plant. *Journal of Agricultural and Food Chemistry*, 45(4), 1027–1031. <https://doi.org/10.1021/jf960637v>
- Rakocevic, M., Costes, E., & Assad, E. D. (2011). Structural and physiological sexual dimorphism estimated from three-dimensional virtual trees of yerba-mate (Ilex paraguariensis) is modified by cultivation environment. *Annals of Applied Biology*, 159(2), 178–191. <https://doi.org/10.1111/j.1744-7348.2011.00484.x>
- Rakocevic, M., Janssens, M., & Schere, R. (2012). Light responses and gender issues in the domestication process of yerba-mate, a subtropical evergreen. In *Evergreens: Types, Ecology and Conservation* (Vol. 2008, Issue 10102, pp. 63–96).
- Rakocevic, M., & Martim, S. F. (2011). Time series in analysis of yerba-mate biennial growth modified by environment. *International Journal of Biometeorology*, 55(2), 161–171. <https://doi.org/10.1007/s00484-010-0322-4>
- Rakocevic, M., Scholz Dos Santos, M. B., & Good Kitzberger, C. S. (2018). 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(2). <https://doi.org/10.1080/15538362.2017.1422448>
- Resende, M. D. V., Sturion, J. A., Carvalho, A. P., Simeão, R. M., & Fernandes, J. (2000). Programa de melhoramento da Erva-Mate coordenado pela EMBRAPA - Resultados da avaliação genética de populações, progênies, indivíduos e clones. *Colombo: Embrapa Florestas*, 43(April), 65.
- Riachi, L. G., Simas, D. L. R., Coelho, G. C., Marcellini, P. S., Ribeiro da Silva, A. J., & Bastos de Maria, C. A. (2018). Effect of light intensity and processing conditions on bioactive compounds in maté extracted from yerba mate (Ilex paraguariensis A. St.-Hil.). *Food Chemistry*, 266. <https://doi.org/10.1016/j.foodchem.2018.06.028>
- Rossa, Ü. B., Angelo, A. C., Mazuchowski, J. Z., Westphalen, D. J., Frizon, C. N. T., & Martins, C. E. N. (2017). Influência Da Luminosidade E Fertilizantes Nos Teores De Metilxantinas E Compostos Fenólicos Em Folhas De Erva-Mate. *Ciência Florestal*, 27(4), 1365–1374. <https://doi.org/10.5902/1980509830217>
- Saldaña, M. D. A., Mohamed, R. S., & Mazzafera, P. (2000). Supercritical carbon

- dioxide extraction of methylxanthines from mate tea leaves. *Brazilian Journal of Chemical Engineering*, 17(3), 251–259. <https://doi.org/10.1590/s0104-66322000000300001>
- Schuhli, G. S. e, Penteado Junior, J. F., & Wendling, I. (2019). *Descritores mínimos em cultivares de espécies florestais: uma contribuição para erva-mate*. 25.
- 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. (2021). 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(July 2020). <https://doi.org/10.1016/j.jfca.2020.103795>
- Sturion, J. A., Stuepp, C. A., & Wendling, I. (2017). Genetic parameters estimates and visual selection for leaves production in Ilex paraguariensis. *Bragantia*, 76(4), 492–500. <https://doi.org/10.1590/1678-4499.2016.419>
- Takeshi Matsunaga, F., Rakocevic, M., & Brancher, J. D. (2014). Modeling the 3D structure and rhythmic growth responses to environment in dioecious yerba-mate. *Ecological Modelling*, 290(C), 34–44. <https://doi.org/10.1016/j.ecolmodel.2013.10.035>
- Tomasi, J. D. C., De Lima, G. G., Wendling, I., Helm, C. V., Hansel, F. A., De Godoy, R. C. B., Grunennvaldt, R. L., De Melo, T. O., Tomazzoli, M. M., & Deschamps, C. (2021). Effects of different drying methods on the chemical, nutritional and colour of yerba mate (Ilex paraguariensis) leaves. *International Journal of Food Engineering*, 17(7), 551–560. <https://doi.org/10.1515/ijfe-2020-0312>
- 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*, 26(29), 30356–30364. <https://doi.org/10.1007/s11356-019-06163-1>
- Tormena, C. D., Pauli, E. D., Marcheafave, G. G., Scheel, G. L., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2020). FT-IR biomarkers of sexual dimorphism in yerba-mate plants: Seasonal and light accessibility effects. *Microchemical Journal*, 158(July), 105329. <https://doi.org/10.1016/j.microc.2020.105329>
- Valduga, E., Freitas, R. J. S. de, Reissmann, C. B., & Nakashima, T. (1997). Caracterização química da folha de Ilex paraguariensis St. Hil. (Erva-mate) e de outras espécies utilizadas na adulteração do mate. *Boletim Do Centro de Pesquisa de Processamento de Alimentos*, 15(1). <https://doi.org/10.5380/cep.v15i1.14033>
- van Dam, R. M., Hu, F. B., & Willett, W. C. (2020). Coffee, Caffeine, and Health. *New England Journal of Medicine*, 383(4), 369–378. <https://doi.org/10.1056/nejmra1816604>
- Wendling, I. (2016). Melhoramento de erva-mate: perspectivas. In: *Embrapa Florestas (Ed.), SEMINÁRIO ERVA-MATE XXI: Modernização No Cultivo e Diversificação Do Uso Da Erva-Mate*, 34–37.
- Wendling, I., Dutra, L. F., & Grossi, F. (2007). Produção e sobrevivência de miniestacas e minicepas de erva-mate cultivadas em sistema semi-hidropônico. *Pesquisa Agropecuária Brasileira*, 42(2), 289–292. <https://doi.org/10.1590/s0100-204x2007000200019>
- Wendling, I., Sturion, J. A., Stuepp, C. A., Reis, C. A. F., Ramalho, M. A. P., & de Resende, M. D. V. (2018). Early selection and classification of yerba mate progenies. *Pesquisa Agropecuária Brasileira*, 53(3), 279–286.

<https://doi.org/10.1590/S0100-204X2018000300002>

Wrege, M. S., Soares, M. T. S., Fritzsos, E., Sousa, V. A. de, Aguiar, A. V. de, Bognola, I. A., & Sousa, L. P. de. (2020). Natural distribution of yerba mate in Brazil in the current and future climatic scenarios. *Agrometeoros*, 28. <https://doi.org/10.31062/agrom.v28.e026795>

Zohra, M., & Fawzia, A. (2014). Hemolytic activity of different herbal extracts used in Algeria. *International Journal of Pharma Sciences and Research*, 5(08), 495–500.

Material Suplementar

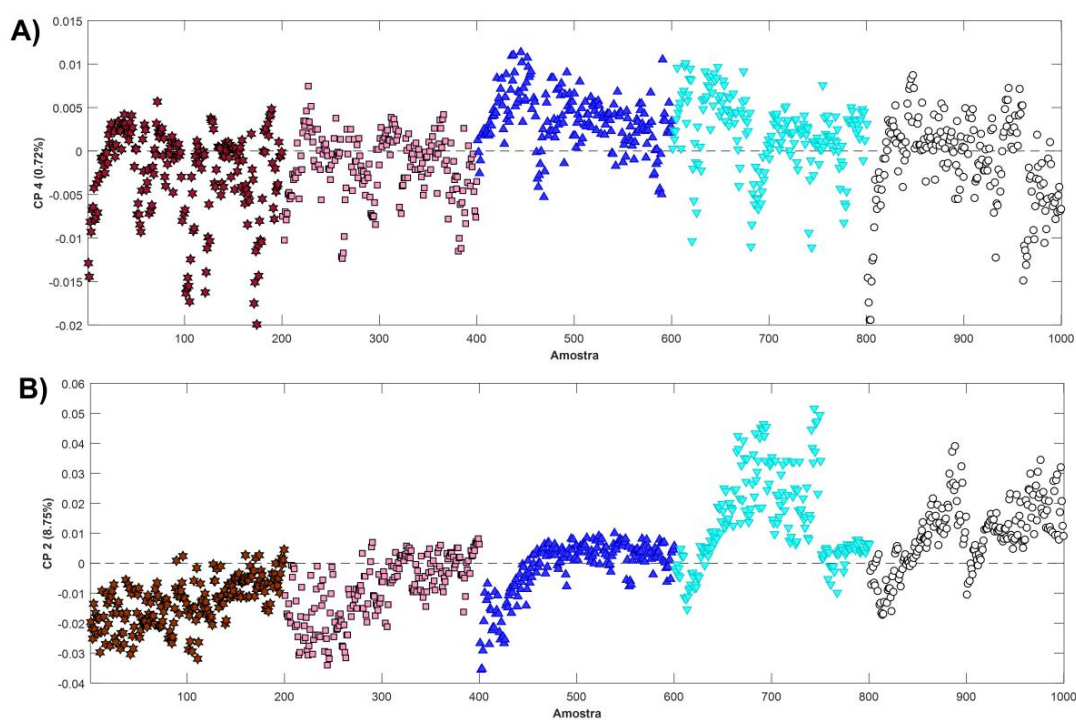


Fig. S1. Gráficos de scores resultantes da análise de componentes principais versus amostras para folhas de erva-mate descafeinadas (a) e com alto teor de cafeína (b) em plantas cultivadas sob cinco diferentes níveis de luz (0%, 40%, 51%, 76%, e 82%), obtidos a partir de 1000 medições NIR na faixa de 900-1700 nm para cada classe. (★) primeiro nível (0%), (■) segundo nível (40%), (▲) terceiro nível (51%), (▼) quarto nível (76%), e (○) quinto nível de sombreamento (82%).

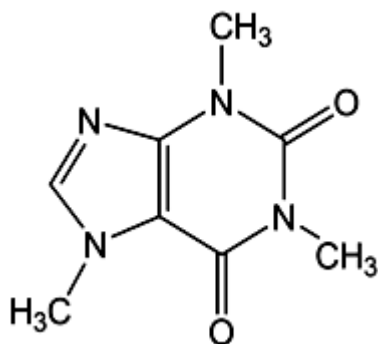


Fig. S2. Estrutura da molécula de cafeína.

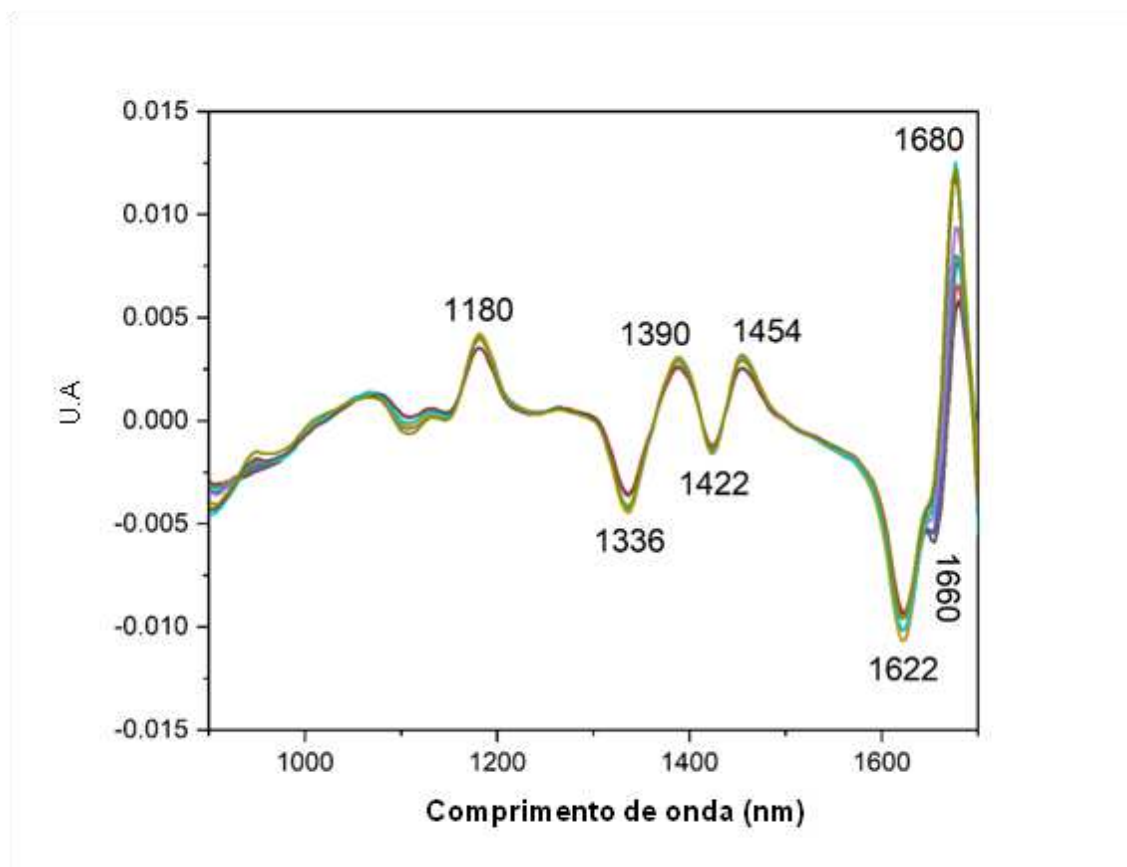


Fig. S3. Dez espectros NIR pré-tratados do padrão de Cafeína com primeira derivada de Savitzky-Golay centrado na média.