

REBECA LOURENÇO DE OLIVEIRA

**SELEÇÃO DE GENÓTIPOS DE ABÓBORA TIPO BUSH E PREDIÇÃO DO TEOR
DE ÓLEO E PERFIL DE ÁCIDOS GRAXOS**

Tese apresentada à Universidade Federal de Viçosa,
como parte das exigências do Programa de Pós-
Graduação em Genética e Melhoramento, para
obtenção do título de *Doctor Scientiae*.

Orientador: Derly José Henriques da Silva

VIÇOSA - MINAS GERAIS

2021

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

O48s
2021
Oliveira, Rebeca Lourenço de, 1987-
Seleção de genótipos de abóbora tipo bush e predição do
teor de óleo e perfil de ácidos graxos / Rebeca Lourenço de
Oliveira. – Viçosa, MG, 2021.
51 f. : il. (algumas color.) ; 29 cm.

Orientador: Derly José Henriques da Silva.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. *Cucurbita moschata*. 2. Melhoramento genético.
3. Espectroscopia de infravermelho próximo. 4. Análise
multivariada. I. Universidade Federal de Viçosa. Departamento
de Agronomia. Programa de Pós-Graduação em Genética e
Melhoramento. II. Título.

CDD 22. ed. 635.62

REBECA LOURENÇO DE OLIVEIRA

**SELEÇÃO DE GENÓTIPOS DE ABÓBORA TIPO BUSH E
PREDIÇÃO DO TEOR DE ÓLEO E PERFIL DE ÁCIDOS GRAXOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 12 de março de 2021

Assentimento:

Rebeca Lourenço de Oliveira

Rebeca Lourenço de Oliveira
Autora

Derly José Henriques da Silva

Derly José Henriques da Silva
Orientador

Aos meus pais João Carlos e Cláudia pelo amor a mim dedicado e pelo apoio constante.

AGRADECIMENTOS

A Deus, pela força nos momentos mais difíceis,

Aos meus pais, João Carlos e Cláudia, pelo apoio, carinho, confiança e por sempre terem acreditado em mim,

As minhas irmãs Letícia e Edivânia, pela amizade, conversas e conselhos,

Aos meus avós: Argemiro (*in memoriam*) e Constantino (*in memoriam*), Zilda e Arlete, aos meus tios: Mara, Luciano, Maninho, Luiza, Darlene e aos meus primos: Angélica, Ana Paula, Carol e Milton, pelo apoio, incentivo, brincadeiras e constante carinho,

A minha afilhada Ana Luiza e a Maria Helena, pelo carinho e por todos os momentos proporcionados de alegria,

Ao meu namorado Pedro, pelo amor, compreensão e ajuda durante todo esse período,

Ao meu orientador Derly José Henriques da Silva, pelos ensinamentos e orientação,

Aos professores Felipe Silva, Carlos Nick, Maicon Nardino e a Lisandra Moura pela disponibilidade e contribuições,

Aos amigos do Núcleo de Estudo em Olericultura (NEO), que sempre me ajudaram, em especial aos amigos de pesquisa em Abóbora, Renata, Cleverson, Ronaldo Gomes e Ronaldo Machado pelas conversas, apoio, sugestões e ajuda com os experimentos,

A Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Genética e Melhoramento pela oportunidade de realização desse curso,

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos,

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001,

E a todos que direta ou indiretamente fizeram parte da realização desse trabalho,

Muito Obrigada!

RESUMO

OLIVEIRA, Rebeca Lourenço, D.Sc., Universidade Federal de Viçosa, março de 2021. **Seleção de genótipos de abóbora tipo Bush e predição do teor de óleo e perfil de ácidos graxos.** Orientador: Derly José Henriques da Silva.

A abóbora (*Cucurbita moschata* Duchesne) é uma olerícola de grande importância socioeconômica e alimentar, com alto valor nutricional de seus frutos e sementes. As sementes da abóbora possuem alto teor de óleo (~ 40%), com cerca de 70% do óleo composto por ácidos graxos insaturados, sendo a maior parte de ácido linoleico e oleico. O incremento da produção de óleo pode ser feito mediante uma maior população de plantas por área. Para isso, plantas de abóbora com hábito de crescimento reduzido podem ser cultivadas em um menor espaçamento, aumentando assim o número de plantas por área. As metodologias convencionais de análises químicas, são demoradas, caras e utilizam reagentes, como alternativa tem a espectroscopia de infravermelho próximo, que além de ser um método rápido possibilita a análise simultânea de diversas características agilizando todo o processo. Em programas de melhoramento genético a seleção de genótipos superiores com as características desejadas é etapa crucial. Essa seleção torna-se mais efetiva quando leva em consideração as várias características de interesse simultaneamente. A utilização do índice de seleção FAI-BLUP (Factor Analysis and Ideotype-Design) direciona o melhorista na seleção de genótipos mais próximo ao ideótipo. Diante do exposto, o trabalho teve como objetivo construir e validar um modelo de calibração para predição do teor de óleo e perfil dos ácidos graxos em genótipos de abóbora utilizando o NIR, além de selecionar genótipos com menor comprimento de rama e alta produtividade de sementes para produção de óleo através do índice de seleção FAI-BLUP.

Palavras-chave: Análise multivariada. *Cucurbita moschata*. FAI-BLUP. NIR.

ABSTRACT

OLIVEIRA, Rebeca Lourenço, D.Sc., Universidade Federal de Viçosa, March, 2021. **Selection of bush type pumpkin genotypes and prediction of oil content and fatty acid profile.** Adviser: Derly José Henriques da Silva.

Pumpkin (*Cucurbita moschata* Duchesne) is a vegetable of great socioeconomic and food importance, with high nutritional value of its fruits and seeds. Pumpkin seeds have a high oil content (~ 40%), with about 70% of the oil composed of unsaturated fatty acids, most of which is linoleic and oleic acid. The increase in oil production can be done through a greater number of plants per area. Therefore, pumpkin plants with reduced growth habits can be grown in a shorter spacing, thus increasing the plant population by area. Conventional chemical analysis methodologies, are time-consuming, expensive and use reagents, as an alternative there is near-infrared spectroscopy, which in addition to be a fast method allows the simultaneous analysis of several characteristics, speeding up the entire process. In breeding programs, the selection of superior genotypes with the desired characteristics is a crucial step. This selection becomes more effective when considers the various characteristics of interest simultaneously. The use of the FAI-BLUP selection index (Factor Analysis and Ideotype-Design) directs the breeder in the selection of genotypes closest to the ideotype. Given the above, the work aimed to build and validate a calibration model to predict oil content and fatty acid profile in pumpkin genotypes using the NIR, in addition to selecting genotypes with shorter internode length and high seed productivity for oil production using the FAI-BLUP selection index.

Keywords: *Cucurbita moschata*. FAI-BLUP. Multivariate analysis. NIR.

SUMÁRIO

INTRODUÇÃO GERAL	8
REFERÊNCIAS	10
ARTIGO 1 - Multi-trait Selection of Pumpkin Genotypes Aimed at Reducing the Growth Habit and Improving Seed Production	12
ABSTRACT	12
1 INTRODUCTION	13
2 MATERIALS AND METHODS	15
2.1 Selection of vegetable material and backcross generations.....	15
2.2 Installation and experimental conditions	16
2.3 Characteristics evaluated	16
2.4 Statistical models	16
2.4.1 Augmented block design.....	16
2.4.2 Spatial model.....	17
2.4.3 Choice of the best-fitted spatial model using the LRT.....	18
2.4.4 Spatial model × ABD.....	18
2.4.5 Accuracy.....	18
2.5 FAI-BLUP selection index	19
2.6 Predicted genetic gain.....	20
2.7 Statistical analysis software.....	20
3 RESULTS AND DISCUSSION	21
4 CONCLUSIONS	29
5 REFERENCES	30
ARTIGO 2 - Determinação do teor de óleo e perfil de ácidos graxos em sementes de abóbora utilizando espectroscopia no infravermelho próximo	34
RESUMO	34
1. INTRODUÇÃO	35
2. MATERIAL E MÉTODOS	36
2.1. Caracterização geral do experimento	36
2.2. Extração do óleo das sementes	36
2.3. Análise de perfil dos ácidos graxos - Cromatografia gasosa.....	37
2.4. Análise espectral - NIR.....	38
2.5. Modelos de regressão multivariada	39
2.6. Predição de novos genótipos	41
3. RESULTADOS E DISCUSSÃO	41
3.1. Análise de referência do teor de óleo e composição de ácidos graxos.....	41
3.2. Calibração e validação dos modelos NIR.....	42
3.3. Predição do teor de óleo e ácidos graxos em amostras desconhecidas	44
4. CONCLUSÕES	46
5. REFERÊNCIAS	47
CONSIDERAÇÕES FINAIS	51

INTRODUÇÃO GERAL

A abóbora (*Cucurbita moschata* Duchesne) pertence à família Cucurbitaceae e ao gênero *Cucurbita*. É originária da região central do México, estendendo-se até a Colômbia e Venezuela (Sazaki et al., 2006). O gênero *Cucurbita* possui 20 pares de cromossomos ($2n=2x=40$) (Paris et al., 2005), e é formado por 24 espécies sendo cinco dessas cultivadas: *C. maxima*, *C. moschata*, *C. pepo*, *C. argyrosperma* e *C. ficifolia* (Nee, 1990).

As plantas de *C. moschata* são anuais, possuem caule herbáceo, rastejante com gavinhas e raízes adventícias que auxiliam na fixação da planta ao solo. Em geral, apresenta hábito de crescimento indeterminado com extensas ramas. É uma planta alógama, com hábito de florescimento monoico, as flores femininas possuem o ovário bem destacado que prenuncia o formato do futuro fruto (Filgueira, 2008).

É uma olerícola de grande importância socioeconômica e alimentar, com alto valor nutricional de seus frutos e sementes, sendo fonte de vitaminas, sais minerais, polissacarídeos, carotenoides entre outros compostos benéficos à saúde (Chen e Huang, 2019).

O uso do óleo das sementes de abóbora para fins alimentares possui aspecto promissor na produção de *C. moschata*. As sementes da abóbora possuem alto teor de óleo, com aproximadamente 40%. Cerca de 70% do óleo é composto por ácidos graxos insaturados, sendo a maior parte de ácido linoleico e oleico (Veronezi e Jorge, 2015; Petkova e Antova, 2015). O óleo é rico em compostos bioativos com propriedades funcionais que promovem benefícios a saúde (Bardaa et al., 2016; Cuco et al., 2019). Em razão dessas características, o óleo das sementes de abóbora é uma excelente alternativa a fontes lipídicas ricas em ácidos graxos saturados.

O aumento da produção de óleo pode ser feito mediante um maior número de plantas por área. No entanto, as cultivares de *C. moschata* possuem hábito de crescimento indeterminado e internódios longos, o que faz com que essas plantas ocupem áreas muito grande (Maynard et al., 2002; Wu et al., 2007). Plantas de abóbora com hábito de crescimento reduzido podem ser cultivadas em um menor espaçamento, aumentando assim a população de plantas por área (Maynard et al., 2002), além de facilitar os tratos culturais e a colheita. Assim, o aumento da produtividade é obtido, pelo fato de que as plantas produzem frutos menores, porém em maior quantidade (Wu et al., 2007), e pelo aumento do número de plantas por área.

A quantificação do teor e perfil do óleo das sementes de abóbora torna-se fundamental para o conhecimento dos genótipos, contudo as técnicas para essa determinação são caras e onerosas. Portanto, o uso de metodologias rápidas, mais baratas e que sejam de fácil execução

na determinação de propriedades químicas de sementes, como o teor de óleo e perfil dos ácidos graxos, torna-se essencial nas etapas de um programa de melhoramento. Como alternativa as técnicas convencionais tem-se a espectroscopia do infravermelho próximo (NIR). Um método que além de ser mais rápido não requer a utilização de reagentes químicos e que possibilita a análise simultânea de diversos parâmetros das amostras (Roque et al., 2017; Li et al., 2020).

Os programas de melhoramento genético buscam a seleção de genótipos superiores que possuem características desejadas. Esta seleção pode ser mais efetiva quando realizada levando em consideração as várias características de interesse simultaneamente, para isso, utiliza-se um índice de seleção. Um índice novo e que vem se mostrando promissor é o índice de seleção FAI-BLUP (Factor Analysis and Ideotype-Design). Este índice se baseia na análise de fatores e considera as correlações genéticas dos dados, direcionando o melhorista na seleção de genótipos mais próximos do ideótipo contemplando múltiplas características e em várias gerações, livre de multicolinearidade e sem precisar atribuir pesos econômicos as características (Rocha et al., 2018).

O objetivo deste trabalho foi construir e validar um modelo de calibração para predição do teor de óleo e perfil dos ácidos graxos em genótipos de abóbora utilizando a espectroscopia no infravermelho próximo (NIR), além de selecionar genótipos com menor comprimento de rama e alta produtividade de sementes para produção de óleo através do índice de seleção FAI-BLUP.

REFERÊNCIAS

BARDAA, S.; HALIMA, N. B.; ALOUI, F.; MANSOUR, R. B.; JABEUR, H.; BOUAZIZ, M.; SAHNOUN, Z. Oil from pumpkin (*Cucurbita pepo* L.) seeds: evaluation of its functional properties on wound healing in rats. **Lipids in health and disease**, v.15, n.73, p. 1-12, 2016.

CHEN, L.; HUANG, G. Antioxidant activities of phosphorylated pumpkin polysaccharide. **International Journal of Biological Macromolecules**, v. 126, p. 743–746, 2019.

CUCO, R. P.; MASSA, T. B.; POSTAUE, N.; CARDOZO-FILHO, L.; DA SILVA, C. Oil extraction from structured bed of pumpkin seeds and peel using compressed propane as solvent. **The Journal of Supercritical Fluids**, 152, 104568, 2019.

FILGUEIRA, F. A. R. Novo manual de olericultura: agrotecnologia moderna na produção e comercialização de hortaliças (3rd ed.). Viçosa, MG: Universidade Federal de Viçosa. 2008.

LI, X.; ZHANG, L.; ZHANG, Y.; WANG, D.; WANG, X.; YU, L.; ZHANG, W.; LI, P. Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils. **Trends in Food Science & Technology**, v. 101, p. 172-181, 2020.

MAYNARD, D.N.; ELMSTROM, G.W.; CARLE, R.B. ‘El Dorado’ and ‘La Estrella’: compact plant tropical pumpkin hybrids. **HortScience**, v.37, p.831–833, 2002.

NEE, M. The domestication of *Cucurbita* (Cucurbitaceae). *Economic Botany*, v.44, p.56-68, 1990.

PARIS, H. S.; BROWN, R. N. The genes of pumpkin and squash. **HortScience**, v. 40, p. 1620-1630, 2005.

PETKOVA, Z. Y.; ANTOVA, G. A. Changes in the composition of pumpkin seeds (*Cucurbita moschata*) during development and maturation. **Grasas y aceites**, v. 66, n. 1, p. 058, 2015.

ROCHA, J.R.A.S.C.; MACHADO, J.C.; CARNEIRO, P.C.S. Multitrait index based on factor analysis and ideotype-design: Proposal and application on elephant grass breeding for bioenergy. **Global Change Biology Bioenergy**, v. 10, p. 52-60, 2018.

ROQUE, J. V.; DIAS, L. A. S.; TEÓFILO, R. F. Multivariate calibration to determine phorbol esters in seeds of *Jatropha curcas* L. using near infrared and ultraviolet spectroscopies. **Journal of the Brazilian Chemical Society**, v. 28, n. 8, p. 1506-1516, 2017.

SASAKI, F. F., DEL AGUILA, J. S., GALLO, C. R., ORTEGA, E. M. M., JACOMINO, A. P., KLUGE, R. A. Alterações fisiológicas, qualitativas e microbiológicas durante o armazenamento de abóbora minimamente processada em diferentes tipos de corte. **Horticultura Brasileira**, v. 24, p. 170-174, 2006.

VERONEZI, C. M. & JORGE, N. Chemical characterization of the lipid fractions of pumpkin seeds. **Nutrition & Food Science**, 45, 164-173, 2015.

WU, T.; ZHOU, J.; ZHANG, Y.; CAO, J. Characterization and inheritance of a bush-type in tropical pumpkin (*Cucurbita moschata* Duchesne). **Scientia Horticulturae**, v. 114, n. 1, p. 1-4, 2007.

ARTIGO 1 - Multi-trait Selection of Pumpkin Genotypes Aimed at Reducing the Growth Habit and Improving Seed Production

Artigo publicado na revista Crop Science em 2021

ABSTRACT

Pumpkin (*Cucurbita moschata* Duchesne) has high food importance due to the high nutritional value of its fruit and seeds. Its seed oil is composed of ~70% unsaturated fatty acids, and its production can be improved by increasing the number of plants per area. Pumpkin plants with reduced growth habits can be grown in shorter spacing and produce smaller fruits, but have a greater quantity of fruits per area. This study aimed to select from F₂, BC_{1:1}, and BC_{1:2} populations pumpkin genotypes with shorter internodes and higher seed productivity for oil production using the factor analysis and ideotype design (FAI-BLUP) selection index. An augmented block design (ABD) was used with four replications. The traditional model of ABD analysis was compared with the spatial model to verify the presence of errors in the local and global trends in the experimental area. There was a significant difference for all traits evaluated in both the ABD and spatial models. Based on the Akaike information criterion, the model with the best fit for all characteristics was the ABD, which was efficient in guaranteeing the independence of the residues and the absence of errors in local and global trends. In general, the characteristics expressed high values of individual heritability and accuracy. Three common factors explained the seven characteristics studied, presenting a variability of 70.98%. Six of the seven characteristics showed desirable gains of up to 63.85%. The FAI-BLUP index allowed the selection of superior pumpkin genotypes with a smaller size and higher seed productivity for oil production.

1 INTRODUCTION

The pumpkin (*Cucurbita moschata* Duchesne) is an allogamous plant belonging to the family Cucurbitaceae and genus *Cucurbita*. All species of this genus are diploid and have 20 pairs of chromosomes ($2n = 2x = 40$). The genus is composed of 24 species, five of which are cultivated: *C. maxima* Duchesne, *C. moschata* Duchesne, *C. pepo* L., *C. argyrosperma* K. Koch, and *C. ficifolia* Bouché (Lira-Saade, 1995). The center of origin of *C. moschata* is Mexico, extending to Colombia and Venezuela (Nee, 1990).

In terms of economics, *C. moschata* is one of the species with the highest socioeconomic importance of the genus *Cucurbita*, which is partly due to the high volume and value of its production. Worldwide, the largest pumpkin producer is China, with a production of just over 8 Tg (FAO, 2018). Brazil produces approximately 418,000 metric tons, with the southeast region being the largest producer, with more than 133,000 metric tons, corresponding to ~32% of the national production (IBGE, 2017).

The pumpkin has great socioeconomic and food importance, mainly for family farming, and constitutes an important role in the basic diet of the Brazilian population (Carvalho, Peixoto, & Ferreira, 2011). It has high nutritional value in its fruit and seeds, with wide use of its products and by-products. The entire plant can be used, from the vegetative parts to the seeds, which makes it attractive for production (Jacobo-Valenzuela et al., 2011).

The oil extracted from pumpkin seeds consists of ~70% unsaturated fatty acids that are rich in bioactive components such as vitamin E, carotenoids, and tocopherols, which protect the oil from oxidation and are important components of antioxidant activity (Veronezi & Jorge, 2015). Pumpkin seed oil has great potential for use in food and industrial industries, making it an excellent substitute for vegetable lipid sources that contain high levels of saturated fatty acids. However, most of its seeds are wasted (Naves, Corrêa, Abreu, & Santos, 2010).

Oil production can be increased by increasing the number of plants per area. However, *C. moschata* cultivars have an indeterminate growth habit and long internodes, which emit branches that extend great distances from the central crown of the plant (Maynard, Elmstrom, & Carle, 2002; Wu, Zhou, Zhang, & Cao, 2007). Pumpkin plants with reduced growth habits can be grown in shorter spacing, which increases the plant population by area and allows greater precocity (Maynard et al., 2002), in addition to facilitating cultural treatments and harvesting. Thus, increased productivity is obtained because the plants produce smaller fruits, but in greater quantity (Wu et al., 2007), and there is an increase in the number of plants per area.

An alternative to decreasing the internode length in *C. moschata* is the introgression of the *Bush* (*Bu*) gene. This gene produces shrub-type plants, with a significant reduction in internode length. When in dominant homozygosity, the *Bu* gene reduces the length of the plant internode (Zhang et al., 2015); however, in heterozygosity, there is a reversal of dominance after flowering of the plants.

The selection of superior genotypes with desired characteristics is of paramount importance for breeding programs, becoming more effective when undertaken after considering the various characteristics of interest simultaneously. For this, a selection index must be used. Individual genotypic averages can be used to increase the gains from the indices, representing a measure of the additive genetic value that constitutes the next generation. Individual genetic values can be estimated using mixed models via the prediction of genotypic values with the best linear unbiased prediction (BLUP) method (Resende, 2009).

The use of an ideotype, a hypothetical model built from a set of desirable agronomic characteristics, provides breeders with a final goal for selection, thus replacing the attempts and errors of gradually increasing the performance of the plant (Van Oijen & Höglind, 2016).

The factor analysis and ideotype design (FAI-BLUP) selection index is based on the analysis of factors and considers the genetic correlations of the data, directing the breeder in the selection of genotypes closest to the ideotype covering multiple characteristics and over several generations. This method is free of multicollinearity and does not require attributing economic weights to the characteristics (Rocha, Machado, & Carneiro, 2018).

The aim of the present study was to select pumpkin genotypes with shorter internode lengths and high seed productivity for oil production using the FAI-BLUP selection index.

2 MATERIALS AND METHODS

2.1 Selection of vegetable material and backcross generations

To obtain the backcross populations, we used the accession BGH-7765 (recurrent parent) from the Vegetable Germplasm Bank at the Federal University of Viçosa. This accession has a high oil content (~44%) and is considered promising for breeding programs aimed at the production of functional pumpkin seed oil. The cultivar ‘Tronco Verde’ (donor parent) was used because it has the *Bush* gene in homozygosity, giving the shrub-like appearance to the plant (i.e., reduced size).

The F₁ hybrid was obtained from the crossing of the parents, BGH-7765 and ‘Tronco Verde’, and later through self-fertilization of the same F₁ plants, the F₂ population was obtained. The F₁ hybrid was backcrossed with the recurrent parent (BGH-7765), resulting in the first generation of backcross (BC_{1:1}). From this generation, plants with shorter branch internode lengths were selected, which were backcrossed with the recurrent parent to generate the second backcross population (BC_{1:2}). The field experiment was composed of all generations: 18 plants of the female parent BGH-7765 (P₁), 18 plants of the male parent ‘Tronco Verde’ (P₂), 50 plants of the F₁ hybrid, 50 plants of the F₂ population, 100 plants of generation BC_{1:1}, and 100 plants of generation BC_{1:2}.

Crosses to obtain the backcross populations (BC_{1:1} and BC_{1:2}) were performed in a controlled way. After the appearance of the first male and female flowers, always during the morning, the flowers were closed; therefore, no contamination occurred from pollinating insects. The selection and closing of the flowers were performed the day before the crosses.

2.2 Installation and experimental conditions

The experiment was performed in the experimental field of the Olericulture sector at the Universidade Federal de Viçosa, Viçosa, Minas Gerais (20°45'14" S, 42°52'53" W). All culture treatments were undertaken in accordance with the recommendations for culturing described by Filgueira (2008).

The experimental design used was Federer's (1956) augmented block design (ABD) with four replications. The controls (BGH-7765 and 'Tronco Verde') were randomly distributed in each of the four blocks, and the 300 regular treatments were randomly designated to all blocks, in 75 regular treatments per block. The spacing used was 3.0 × 4.0 m between the plants and rows.

2.3 Characteristics evaluated

The following characteristics were evaluated: average internode length before flowering (ILB) and average internode length after flowering (ILA), in centimeters, measuring three internodes before and after the first flower, respectively; average internode diameter after flowering (IDA), in millimeters, measuring three internodes after the first flower using a digital caliper; number of fruits per plant (NF); average fruit flesh thickness (FFT), in centimeters, obtained by measuring the four sides of the pulp with a graduated ruler; total seed mass per fruit (SM), in grams; and average seed thickness (ST), in millimeters, measured using a digital caliper.

2.4 Statistical models

2.4.1 Augmented block design

The genotypic means were generated by the restricted maximum likelihood–BLUP method, proposed by Henderson (1975) and Patterson and Thompson (1971), which uses a mixed linear model methodology. The statistical model used to design Federer's (1956) ABD was as follows:

$$\mathbf{y} = \mathbf{Xm} + \mathbf{Zg} + \mathbf{Wb} + \boldsymbol{\varepsilon}$$

where \mathbf{y} is the phenotypic data vector arranged in line and columns; \mathbf{m} is the fixed effect vector (i.e., the general mean), with the incidence matrix \mathbf{X} ; g is the genotypic random effect with the incidence matrix \mathbf{Z} , $g \sim \text{NID}(0, \sigma_g^2)$, where σ_g^2 is the genotypic variance; \mathbf{b} is the vector of the random effects of blocks with the incidence matrix \mathbf{W} , $\mathbf{b} \sim \text{NID}(0, \sigma_b^2)$, where σ_b^2 is the block variance; and $\boldsymbol{\varepsilon}$ is the residue vector (assumed as random), $\boldsymbol{\varepsilon} \sim \text{NID}(0, \sigma_\varepsilon^2)$, where σ_ε^2 is the residual variance.

2.4.2 Spatial model

For the spatial analysis of the data, the following complete spatial model was used:

$$\mathbf{y} = \mathbf{Xm} + \mathbf{Zg} + \mathbf{Wb} + \text{Sc} + \text{Sr} + \boldsymbol{\xi} + \boldsymbol{\eta}$$

where Sc and Sr include the random parts of the cubic splines indexed to the column and line, respectively (Verbyla, Cullis, Kenward, & Welham, 1999). These two terms represent the global trends in the columns and rows, respectively, where $c \sim \text{N}(0, \sigma_{\text{spl col}}^2)$ and $\sigma_{\text{spl col}}^2$ is the variance in the columns, and $l \sim \text{N}(0, \sigma_{\text{spl row}}^2)$, where $\sigma_{\text{spl row}}^2$ is the variance in the rows. $\boldsymbol{\xi}$ is a spatially dependent random error vector in columns and rows. $\boldsymbol{\xi} \sim \text{N}(0, \mathbf{R})$, the matrix $\mathbf{R} = \sigma_\xi^2 [\text{AR1}(\rho_c) \otimes \text{AR1}(\rho_r)]$, \otimes represents the Kronecker product of first order two-dimensional separable autoregressive processes, σ_ξ^2 is the variance of the dependent spatial residue, and ρ_c and ρ_r represent the autoregressive correlation of columns and rows, respectively. $\boldsymbol{\eta}$ is a random vector of independent errors. This term is also called the nugget effect, which represents the

measurement error of each parcel or local trend. $\eta \sim N(0, \sigma_\eta^2)$, where σ_η^2 is the variance of the independent error.

The spatial models were tested for the evaluated characteristics ($p < .05$) based on the likelihood ratio test (LRT). For all effects, only those models whose effects were significant were used for comparison with the traditional model of analysis in the ABD.

2.4.3 Choice of the best-fitted spatial model using the LRT

The random effects of the models were tested using the LRT (Wilks, 1938). The LRT estimate was compared with the tabulated χ^2 value, with a degree of freedom and a significance level of 1 and 5%:

$$\text{LRT} = 2\log(\text{Lc}) - 2\log(\text{Lr})$$

where $\log(\text{Lc})$ is the maximum likelihood function for the complete model and $\log(\text{Lr})$ is the maximum likelihood function for the reduced model.

2.4.4 Spatial model \times ABD

The spatial and ABD models were compared using Akaike's (1974) information criterion (AIC), according to the following formula:

$$\text{AIC} = -2\log(L) + 2p,$$

where $\log(L)$ is the logarithm of the residual likelihood and p is the number of estimated parameters.

The strength of the evidence provided by the difference in the AIC value for choosing the best model should be greater than two (Cavanaugh & Neath, 2019).

2.4.5 Accuracy

The accuracy ($r_{\hat{g}g}$) was estimated using the following formula:

$$r_{\hat{g}g} = \sqrt{1 - \frac{PEV}{\hat{\sigma}_g^2}}$$

where PEV is the prediction error variance for the genotypic values and $\hat{\sigma}_g^2$ is the estimated genotypic variation between genetic treatments.

2.5 FAI-BLUP selection index

The selection index was based on the factor analysis and the genotype–ideotype distance in a multitrait manner. Ideotypes are defined based on the desirable and undesirable meaning of the characteristics according to the objectives of the breeding program in question.

The distances between genotypes and ideotypes were calculated. The genotypes with the shortest distance compared with the ideotype of interest were ranked and estimated and were then converted into spatial probability according to the following formula:

$$P_{ij} = \frac{\frac{1}{d_{ij}}}{\sum_{i=1; j=1}^{i=n; j=m} \frac{1}{d_{ij}}}$$

where P_{ij} is the probability of the i th genotype ($i = 1, 2, \dots, n$) to be like the j th ideotype ($j = 1, 2, \dots, m$) and d_{ij} is the genotype–ideotype distance from the i th genotype to the j th ideotype based on the standardized mean Euclidean distance.

The desirable ideotype used in the FAI-BLUP index is shown in Table 1.

TABLE 1 Pumpkin ideotype with smaller size and high oil productivity using the factor analysis and ideotype design (FAI-BLUP) index

Trait	Ideotype
ILB	Min.
ILA	Min.
IDA	Min.
NF	Max.
FFT	Min.
SM	Max.
ST	Max.

Note. ILB, average internode length before flowering; ILA, average internode length after flowering; IDA, average internode diameter after flowering; NF, number of fruits per plant; FFT, average fruit flesh thickness; SM, total seed mass per fruit; ST, average seed thickness; Min., lowest genotypic value; Max., highest genotypic value.

2.6 Predicted genetic gain

From the genotypes selected with the FAI-BLUP index, the gains for each characteristic were predicted, according to the following equation:

$$GG\% = \frac{\bar{X}_m - \bar{X}_o}{\bar{X}_o} \times 100$$

where GG% is the gain in selecting the best genotype of the population for the variable, \bar{X}_m is the BLUP average of the selected genotype of the population for the variable, and \bar{X}_o is the general BLUP average of all genotypes for the variable.

2.7 Statistical analysis software

The data were analyzed using a mixed model methodology. The ASReml software (Gilmour, Gogel, Cullis, Welham, & Thompson, 2015) was used for individual analysis and LRT. The stages of the factor analysis, principal components, and genotype–ideotype distance were calculated using the FAI-BLUP index routine in R software (R Core Team, 2019), according to the routine described by Rocha et al. (2018).

3 RESULTS AND DISCUSSION

The results from the LRT showed that the genotypes had a significant difference ($p \leq .01$) for all characteristics evaluated in both the ABD and spatial models. Therefore, there was high variability between the genotypes and the possibility of obtaining genetic gains via the selection of superior genotypes (Table 2). Several authors have previously found significant differences in most of the characteristics studied in the present study (Du, Sun, Li, Zhou, & Li, 2011; Kiramana & Isutsa, 2018; Oliveira et al., 2016).

There was no significant difference between the blocks for all variables studied, except for ILB and NF; therefore, the blocks did not influence the values obtained by these variables. Regarding the column splines, there was also no significant difference; thus, there was no global trend. For line splines and units, only the NF characteristics showed a significant difference. For the autoregressive, all evaluated characteristics showed significant differences ($p \leq .01$).

TABLE 2 Significance test for the spatial and augmented block models (ABDs) using the likelihood ratio test (LRT)

Characteristic	Spatial model				ABD model		
	Spl column	Spl row	Units	AR	Gen	Gen	Block
ILB	1.81	1.81	1.81	74.25**	106.58**	113.85**	6.80**
ILA	0.00	0.05	0.00	62.88**	97.82**	99.99**	1.35
IDA	0.00	0.00	3.48	23.00**	26.10**	23.47**	0.00
NF	0.00	4.65*	13.49**	54.47**	23.70**	26.54**	36.76**
FFT	0.08	0.00	0.00	30.59**	29.19**	29.98**	0.00
SM	0.00	3.81	0.02	26.53**	34.83**	30.89**	0.00
ST	-1.04	-0.55	0.30	12.86**	9.50**	8.31**	0.00

Note. ILB, average internode length before flowering; ILA, average internode length after flowering; IDA, average internode diameter after flowering; NF, number of fruits per plant; FFT, average fruit flesh thickness; SM, total seed mass per fruit; ST, average seed thickness. The parameters related to the spatial and ABD models include column splines (Spl column), row splines (Spl row), independent errors (Units), autoregressive-dependent errors (AR), and genotype effects (Gen).

*Significant at the .05 probability level. **Significant at the .01 probability level.

The choice of the model that best fit the data was based on the AIC, where lower AIC values reflect a better adjustment of the model (Table 3). The ILB, ILA, NF, and FFT showed lower

AIC values for the ABD model compared with the spatial model, indicating that ABD had the best model fit. Although the other characteristics (IDA, SM, and ST) had lower AIC values for the spatial model, these values were not greater than two units in relation to the ABD (Cavanaugh & Neath, 2019). Therefore, the ABD model was chosen as the best fit because it was less parameterized.

Accuracy is an important parameter for determining the quality of the experiment because in addition to the residual variation and the number of repetitions, it also depends on the proportion of genetic and residual variations associated with the evaluated variables (Resende & Duarte, 2007). All characteristics showed a high accuracy value (>0.63), with ILB, ILA, SM, and ST having higher values for the ABD model, whereas IDA, NF, and FFT had higher values for the spatial model (Table 3). Based on the methodology described by Resende and Duarte (2007), these accuracy estimates were classified as moderate (0.50–0.65) to very high (0.90–0.99). These results show the existence of population genetic variance for the evaluated characteristics, experimental quality, and reliability of the results obtained.

TABLE 3 Choice of the best model based on the Akaike information criterion (AIC), individual heritability (h^2), and accuracy ($r_{\hat{g}g}$) of the augmented block design (ABD) and spatial models

Characteristic	AIC		h^2		$r_{\hat{g}g}$	
	Spatial	ABD	Spatial	ABD	Spatial	ABD
ILB	1,386.88	1,378.28	.91 (.02)	.90 (.03)	0.91	0.96
ILA	1,333.41	1,332.52	.86 (.04)	.86 (.04)	0.89	0.93
IDA	1,074.70	1,075.22	.65 (.09)	.63 (.10)	0.77	0.76
NF	1,200.35	1,192.63	.61 (.10)	.63 (.10)	0.82	0.81
FFT	632.06	630.97	.71 (.08)	.72 (.08)	0.85	0.81
SM	2,150.20	2,151.83	.67 (.08)	.63 (.10)	0.78	0.80
ST	-267.80	-268.88	.44 (.15)	.45 (.15)	0.63	0.64

Note. ILB, average internode length before flowering; ILA, average internode length after flowering; IDA, average internode diameter after flowering; NF, number of fruits per plant; FFT, average fruit flesh thickness; SM, total seed mass per fruit; ST, average seed thickness. Standard errors in parentheses.

Individual heritability (Table 3) showed high values (>0.63) for all evaluated traits, except for ST that presented a mean value. These results demonstrate a low environmental influence and high possibility of success in the selection of pumpkin genotypes for the population in the present study.

Lower heritability values for internode length and mean FFT were observed by Srikanth, Bharad, Thulasiram & Potduke, 2017 using 23 pumpkin genotypes. In contrast, close heritability values for NF and FFT were found in a study of pumpkin lines (Bezerra Neto et al., 2006) and for total seed mass in seven pumpkin genotypes (Blank, Silva, Matos, Carvalho Filho, & Silva-Mann, 2013).

Regarding the factor analysis, eigenvalues and accumulated variances were obtained for the seven principal components obtained from the genotypic correlation matrix between the variables (Table 4). The first three eigenvalues explained 70.98% of the total variance between genotypes. According to Kaiser's (1958) criterion, the number of factors is defined by the number of eigenvalues equal to or greater than the unit. The group of variables in each factor has a high correlation with each other and a low correlation with the variables of the other factors (Cruz, Carneiro, & Regazzi, 2014). Thus, three factors were used in the FAI-BLUP index to select the genotypes based on the variables analyzed (Table 4).

TABLE 4 Eigenvalues and accumulated variance explained by the principal components (PCs) for the seven characteristics evaluated in the pumpkin genotypes

Principal component	Eigenvalue	Accumulated variance (%)
PC1	2.54	36.23
PC2	1.32	55.08
PC3	1.11	70.98
PC4	0.79	82.28
PC5	0.70	92.31
PC6	0.44	98.66
PC7	0.09	100.00

Based on the factorial loads, which indicate the correlation of each variable with a certain factor, after varimax rotation, the values of the variables that most correlated with each factor were observed (Table 5, highlighted in bold). These factors were defined by the magnitude and signs of the factor load. The signs of the factor load reflect the direction of selection for a certain variable based on the objective of the breeding program (Ferreira et al., 2010). Therefore, the factorial loads of all characteristics showed agreement with the purpose of the breeding program of the present study that is, to reduce the size of the plant and increase the total mass of seeds for oil production.

The first factor, which explained 36.23% of the variance, grouped the characteristics related to the branch (ILB, ILA, and IDA). The second factor presented 18.85% of the variance and consisted of the seed variables (SM and ST), whereas the third factor, with 15.90% of the variance, was composed of the variables related to the fruit (NF and FFT). The variables grouped in the same factor have a greater correlation with each other than with variables of different factors (Cruz et al., 2014; Rocha et al., 2018).

TABLE 5 Factor loads after varimax rotation and commonality for all the variables evaluated in pumpkin genotypes based on the factor analysis and ideotype design (FAI-BLUP) index

Variable	Factor			Commonality
	1	2	3	
ILB	-0.94	-0.10	0.14	0.92
ILA	-0.94	-0.03	0.08	0.90
IDA	-0.53	-0.39	-0.07	0.44
NF	-0.20	-0.17	0.68	0.53
FFT	-0.07	-0.09	-0.83	0.71
SM	-0.09	0.87	0.08	0.77
ST	-0.11	0.84	0.01	0.71
Variance explained, %	36.23	18.85	15.90	

Note. ILB, average internode length before flowering; ILA, average internode length after flowering; IDA, average internode diameter after flowering; NF, number of fruits per plant; FFT, average fruit flesh thickness; SM, total seed mass per fruit; ST, average seed thickness. In bolding, are the values of the variables that most correlated with each factor.

The three common factors, generated by the factor analysis, presented values of commonality ranging from 0.44 to 0.92 (Table 5). The variability between the studied genotypes expressed by the seven variables was described by these three common factors. In the analysis of factors, the variation in the characteristics studied is explained by the common factors, of which the proportion is given by commonality (Cruz et al., 2014).

The commonalities, which represent the proportion of the variance of the standardized character, were values ≥ 0.71 , except for NF and IDA (0.53 and 0.44, respectively). Most of the variables had high commonality values, indicating that the factors had a good association with the evaluated variables (Hair, Black, Babin, Anderson, & Tatham, 2009). The IDA variable had a low commonality value (0.44), indicating a low relationship between the factor and the observed characteristic and, consequently, little practical importance. Therefore, the information obtained from this characteristic is unreliable.

The predicted genotype gains obtained from the selection of the genotypes for all traits evaluated simultaneously are shown in Table 6. Six of the seven characteristics showed gains in a desirable direction, except for NF. These results show the efficiency of the index used to provide gains for the set of evaluated traits.

TABLE 6 Estimates of genotypic gain predicted for all the characteristics evaluated in pumpkin genotypes using the factor analysis and ideotype design (FAI-BLUP) selection index

Factor	Characteristic	Genotypic gain predicted (%)
Branch	ILB	-63.85
	ILA	-55.20
Seed	IDA	-11.02
	SM	4.54
	ST	0.62
Fruit	NF	-11.54
	FFT	-12.58

Note. ILB, average internode length before flowering; ILA, average internode length after flowering; IDA, average internode diameter after flowering; NF, number of fruits per plant; FFT, average fruit flesh thickness; SM, total seed mass per fruit; ST, average seed thickness.

In a pumpkin breeding program aimed at plants with reduced growth habits and increased seed productivity for oil extraction, gains in those characteristics that provide an increase in averages are desirable. However, for the other characteristics, the aim is to decrease their averages. For instance, the internode length should be decreased so that the spacing between the plants is reduced to increase productivity per area, whereas the FT is reduced to produce a larger internal cavity to increase the seed production per fruit and, consequently, increase oil production. A previous study showed that FT is negatively correlated with the internal cavity of the fruit and with the seed mass (Nachbar & Souza, 2017).

Therefore, gains with a negative sign are required for selection in the desirable direction for ILB, ILA, IDA, and FFT (Table 6). The selection must be consistent with the objectives of the program to obtain improved simultaneous gains (Ferreira et al., 2005).

A total of 25 genotypes with the highest potential were selected for the evaluated characteristics, based on the FAI-BLUP index, considering a selection intensity of 10% (Table 7).

Average FFT is important for increasing seed production. Increasing the FT may reduce the size of the internal cavity of the fruit (Blank et al., 2013; Laurindo, Laurindo, Delazari, Carneiro, & Silva, 2017; Pereira, Silva, Almeida, Silva, & Queiroga, 2012). As one of the objectives of the present study was to increase seed production for oil, it is important to have fruits with a smaller FT, which consequently will have a larger internal cavity that has more space for seed production.

The selected genotypes showed shorter internode lengths (ILB and ILA) compared with the total population average, which are of great potential to continue the breeding programs aimed at shortening the branch. Increasing the plant population by area results in higher productivity because even if plants with shorter stem lengths produce smaller fruits than those of vine plants, they can produce a higher quantity of fruit per area (Wu et al., 2007). When interpreting the relationship between ILB and ILA in the original and selected population, we observed that the selected plant population could be doubled in relation to the original population data.

TABLE 7 Genotypic means of the 25 pumpkin genotypes selected by the factor analysis and ideotype design (FAI-BLUP) index aimed at the selection of superior genotypes

Order	Genotype	ILB	ILA	IDA	NF	FFT	SM	ST
		— cm —		mm		cm	g	mm
1	BC _{1:2} _164	5.59	6.67	10.06	10.08	2.68	63.49	1.92
2	F ₂ _48	3.97	4.11	10.94	4.92	3.00	61.58	1.74
3	BC _{1:1} _163	4.20	4.81	8.67	5.69	3.09	54.32	1.62
4	BC _{1:2} _38	5.97	6.54	11.19	4.86	2.79	67.70	1.77
5	F ₂ _221	2.35	3.36	9.49	4.23	2.29	27.04	1.57
6	F ₂ _40	4.99	5.40	13.21	5.59	4.10	70.80	1.85
7	F ₂ _143	3.35	7.58	9.99	3.46	3.45	59.41	1.81
8	F ₂ _237	1.59	3.12	9.49	2.48	3.04	32.59	1.60
9	BC _{1:2} _10	5.52	6.71	10.75	8.58	3.20	50.87	1.84
10	F ₂ _17	3.35	5.55	11.19	4.19	3.77	49.10	1.92
11	F ₂ _189	5.12	7.70	10.37	4.96	2.93	56.29	1.82
12	F ₂ _313	3.27	6.83	8.67	3.50	2.73	44.04	1.62
13	F ₂ _154	7.13	7.46	11.26	7.05	3.27	74.49	1.61
14	BC _{1:2} _102	7.16	9.15	9.55	7.05	3.49	64.70	1.78
15	BC _{1:2} _55	4.59	6.41	11.89	12.97	2.70	46.57	1.59
16	BC _{1:1} _279	8.69	10.64	11.95	3.21	3.20	82.30	1.88
17	BC _{1:1} _84	4.59	6.99	10.94	7.12	2.57	42.46	1.68
18	BC _{1:2} _9	3.35	5.55	12.01	4.19	3.56	60.36	1.48
19	F ₂ _331	4.68	5.73	11.45	2.48	3.22	42.61	1.81
20	BC _{1:1} _133	5.61	8.00	8.54	15.10	1.98	34.92	1.34
21	F ₂ _180	1.59	3.12	9.05	4.68	4.58	30.92	1.44
22	BC _{1:2} _58	5.31	5.97	10.82	7.78	3.36	44.53	1.55
23	BC _{1:2} _19	2.74	5.55	9.49	3.46	3.13	22.00	1.62
24	BC _{1:2} _136	7.13	7.16	11.45	7.78	2.79	45.51	1.75
25	F ₂ _44	4.28	6.41	9.81	5.66	3.90	32.14	1.71
Selected avg.		4.64	6.26	10.49	6.04	3.15	50.43	1.69
Original avg.		13.07	13.83	11.79	6.83	3.61	48.24	1.68

Note. ILB, average internode length before flowering; ILA, average internode length after flowering; IDA, average internode diameter after flowering; NF, number of fruits per plant; FFT, average fruit flesh thickness; SM, total seed mass per fruit; and ST, average seed thickness.

Although the predicted gain of selection for NF was negative (Table 6), the average of the selected genotypes was not much lower than the average of the original population (6.04 and 6.83, respectively) (Table 7). This small loss can be the increase of plants per area that will happen occurring due to the shortening of the internode of these plants. However, the BC_{1:2}_164,

BC_{1:2_55}, and BC_{1:1_133} genotypes stood out, with production that was higher than 10 fruits per plant.

For SM, 13 genotypes showed higher values than the average; therefore, the selection was satisfactory for this characteristic. The BC_{1:1_279} genotype had the highest SM per fruit (82.3 g). According to Gomes et al. (2020), there is a high positive correlation between the characteristics of seeds and oil productivity. Therefore, with greater production of seeds per fruit, seed oil productivity is increased.

The FAI-BLUP selection index was efficient in the selection of superior pumpkin genotypes based on the evaluated characteristics. Although only recently proposed, this index has already been used efficiently in the evaluation and selection of sorghum hybrids for increased biomass (Oliveira et al., 2019; Silva et al., 2018), selection of soybeans [*Glycine max* (L.) Merr.] for biodiesel production (Woyann et al., 2019), and selection of superior bean (*Phaseolus vulgaris* L.) progenies (Rocha et al., 2019).

4 CONCLUSIONS

The studied characteristics showed high heritability and accuracy and had a great possibility of success for the selection of pumpkin genotypes from the study population, in addition to good experimental precision. The factorial loads of all characteristics showed agreement with the objectives of the breeding program. The ILB, ILA, IDA, FFT, SM, and ST showed desirable genotypic gains. The FAI-BLUP index allowed the selection of superior pumpkin genotypes with a smaller size and higher seed productivity, which are promising for the production of seed oil.

5 REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.
- Blank, A. F., Silva, T. B., Matos, M. L., Carvalho Filho, J. L. S. & Silva-Mann, R. (2013). Parâmetros genotípicos, fenotípicos e ambientais para caracteres morfológicos e agronômicos em abóbora. *Horticultura Brasileira*, 31, 106-111.
- Bezerra Neto, F. V., Leal, N. R., Costa, F. R., Gonçalves, G. M., Amaral Júnior, A. T., Vasconcellos, H. O. & Mello, M. (2006). Análise biométrica de linhagens de abóbora. *Horticultura Brasileira*, 24, 378-380.
- Carvalho, P. G. B., Peixoto, A. A. P. & Ferreira, M. A. J. F. (2011). *Caracterização de abóboras quanto aos teores de carotenoides totais, alfa e beta-caroteno* (Boletim de Pesquisa e Desenvolvimento). Brasília: Embrapa Hortaliças.
- Cavanaugh, J. E & Neath, A. A. (2019). The Akaike information criterion: background, derivation, properties, application, interpretation and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, 11, 1460–1471.
- Cruz, C. D., Carneiro, P. C. S. & Regazzi, A. J. (2014). *Modelos Biométricos aplicados ao melhoramento genético* (3rd ed. 2nd vol). Viçosa, MG: UFV.
- Du, X., Sun, Y., Li, X., Zhou, J. & Li, X. (2011). Genetic divergence among inbred lines in *Cucurbita moschata* from China. *Scientia Horticulturae*, 127, 207-213.
- Food and Agriculture Organization. (2018). *Crops Production, 2018*. Rome: FAO.
- Federer, W. T. (1956). Augmented (or hoonuiaku) designs. *Hawaiian Planter's Record*, 40, 191-208.
- Ferreira, F. M., Neves, L. G., Bruckner, C. H., Vianna, A. P., Cruz, C. D. & Barelli, M. A. A. (2010). Formação de supercaracteres para seleção de famílias de maracujazeiro amarelo. *Acta Scientiarum Agronomy*, 32, 247-254.
- Ferreira, A., Cecon, P. R., Cruz, C. D., Ferrão, R. G., Silva, M. F., Fonseca, A. F. A. & Ferrão, M. A. G. (2005). Seleção simultânea de *Coffea canephora* por meio da combinação de análise de fatores e índices de seleção. *Pesquisa Agropecuária Brasileira*, 40, 1189-1195.
- Filgueira, F. A. R. (2008). *Novo manual de olericultura: agrotecnologia moderna na produção e comercialização de hortaliças* (3rd ed.). Viçosa, MG: Universidade Federal de Viçosa.
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. & Thompson, R. (2015). ASReml user guide release 4.1 structural specification. *Hemel hempstead: VSN International ltd*.

- Gomes, R. S., Machado Júnior, R., Almeida, C. F., Chagas, R. R., Oliveira, R. L., Delazari, F. T. & Silva, D. J. H. (2020). Brazilian germplasm of winter squash (*Cucurbita moschata* D.) displays vast genetic variability, allowing identification of promising genotypes for agro-morphological traits. *Plos One*, 15, e0230546.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. (2009). *Análise multivariada de dados* (6th ed.). Porto Alegre: Bookman.
- Henderson, C. H. (1975). Best linear estimation and prediction under a selection model. *Biometrics*, 31, 423-447.
- Instituto Brasileiro de Geografia e Estatística. (2017). *Censo Agropecuário 2017: resultados definitivos*. Rio de Janeiro: IBGE.
- Jacobo-Valenzuela, N., Zazueta-Morales, J. J., Gallegos-Infante, J. A., Aguilar-Gutierrez, I. L., Camacho-Hernandez, N. E., Rocha-Guzman, E. & Gonzalez-Laredo, R. F. (2011). Chemical and physicochemical characterization of winter squash (*Cucurbita moschata* D.). *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 39, 34-40.
- Kaiser, H. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Kiramana, J. K. & Isutsa, D. K. (2018). Multivariate analysis techniques reveal significant morphogenetic variability in pumpkin landraces in Kenya. *International Journal of Biosciences*, 13, 298-319.
- Laurindo, R. D. F., Laurindo, B. S., Delazari, F. T., Carneiro, P. C. S & Silva, D. J. H. (2017). Potencial de híbridos e populações segregantes de abóbora para teor de óleo nas sementes e plantas com crescimento do tipo moita. *Revista Ceres*, 64, 582-591.
- Lira-Saade, R. L. (1995). *Estudios taxonomicos ecogeograficos de las Cucurbitaceae latinoamericanas de importancia economica*. Rome, Itália: International Plant Genetic Resources Institute.
- Maynard, D. N., Elmstrom, G. W. & Carle, R. B. (2002). ‘EI Dorado’ and ‘La Estrella’: compact plant tropical pumpkin hybrids. *HortScience*, 37, 831-833.
- Nachbar, L. A. & Souza, S. A. M. (2017). Correlações entre caracteres morfoagronômicos de frutos de genótipos de *Cucurbita* spp. *Revista de Ciências Agroambientais*, 15, 175-180.
- Naves, L. P., Corrêa, A. D., Abreu, C. M. P. & Santos, C. D. (2010). Nutrientes e propriedades funcionais em sementes de abóbora (*Cucurbita maxima*) submetidas a diferentes processamentos. *Ciência e Tecnologia de Alimentos*, 30, 185-190.
- Nee, M. (1990). The domestication of *Cucurbita* (Cucurbitaceae). *Economic Botany*, 44, 56-68.

- Oliveira, R. L., Gonçalves, L. S. A., Rodrigues, R., Baba, V. Y., Sudré, C. P., Santos, M. H. & Aranha, F. M. (2016). Genetic divergence among pumpkin landraces. *Semina: Ciências Agrárias*, 37, 547-556.
- Oliveira, I. C. M., Marçal, T. S., Bernadino, K. C., Ribeiro, P. C. O., Parrella, R. A. C., Carneiro, P. C. S., ... Carneiro, J. E. S. (2019). Combining ability of biomass sorghum lines for agroindustrial characters and multitrait selection of photosensitive hybrids for energy cogeneration. *Crop Science*, 59, 1554-1566.
- Pereira, A. M., Silva, G. D., Almeida, R. R. P., Silva, A. B. & Queiroga, R. C. F. (2012). Frutificação de abóbora Tetsukabuto sobe aplicação de doses de 2,4-D na época seca em Pombal-PB. *Revista Verde de Agroecologia e Desenvolvimento Sustentável*, 7, 38-43.
- Patterson, H. D. & Thompson, R. (1971). Recovery of inter-block information when blocks sizes are unequal. *Biometrika*, 58, 545–554.
- R Core Team. (2019). R: A language and environment for statistical computing [Software]. Vienna, Austria: R Foundation for Statistical Computing.
- Resende, M. D. V. (2009). *Genética biométrica e estatística no melhoramento de plantas perenes*. Brasília: Embrapa Informação Tecnológica.
- Resende, M. D. V. & Duarte, J. B. (2007). Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesquisa Agropecuária Tropical*, 37, 182-194.
- Rocha, J. R. A. S. C., Nunes, K. V., Carneiro, A. L. N., Marçal, T. D. S., Salvador, F. V., Carneiro, P. C. S. & Carneiro, J. E. S. (2019). Selection of superior inbred progenies toward the common bean ideotype. *Agronomy Journal*, 11, 1181–1189.
- Rocha, J. R. A. S. C., Machado, J. C. & Carneiro, P. C. S. (2018). Multitrait index based on factor analysis and ideotype-design: Proposal and application on elephant grass breeding for bioenergy. *Global Change Biology Bioenergy*, 10, 52-60.
- Silva, M. J., Carneiro, P. C. S., Carneiro, J. E. S., Damasceno, C. M. B., Parrella, N. N. L. D., Pastina, M. M., ... Parrella, R. A. C. (2018). Evaluation of the potential of lines and hybrids of biomass sorghum. *Industrial Crops and Products*, 125, 379–385.
- Srikanth, M., Bharad, S. G., Thulasiram, L. B. & Potdukhe, N. R. (2017). Studies on genetic variability, heritability and genetic advance in pumpkin (*Cucurbita moschata* Duch ex Poir.). *International Journal of Current Microbiology and Applied Sciences*, 6, 1416-1422.
- Van Oijen, M. & Höglind, M. (2016). Toward a Bayesian procedure for using process-based models in plant breeding, with application to ideotype design. *Euphytica*, 207, 627-643.

- Verbyla, A. P., Cullis, B. R., Kenward, M. G. & Welham, S. J. (1999). The analyses of designed experiments and longitudinal data using smoothing splines. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48, 269-311.
- Veronezi, C. M. & Jorge, N. (2015). Chemical characterization of the lipid fractions of pumpkin seeds. *Nutrition & Food Science*, 45, 164-173.
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9, 60- 62.
- Woyann, L. G., Meira, D., Zdziarski, A. D., Matei, G., Mililoi, A. S., Rosa, A. C., ... Benin, G. (2019). Multiple-trait selection of soybean for biodiesel production in Brazil. *Industrial Crops and Products*, 140, 111721.
- Wu, T., Zhou, J., Zhang, Y. & Cao, J. (2007). Characterization and inheritance of a bush-type in tropical pumpkin (*Cucurbita moschata* Duchesne). *Scientia Horticulturae*, 114, 1-4.
- Zhang, G., Ren, Y., Sun, H., Guo, S., Zhang, F., Zhang, ... Li, H. (2015). A high-density genetic map for anchoring genome sequences and identifying QTLs associated with dwarf vine in pumpkin (*Cucurbita maxima* Duch.). *BMC genomics*, 16, 1-13.

ARTIGO 2 - Determinação do teor de óleo e perfil de ácidos graxos em sementes de abóbora utilizando espectroscopia no infravermelho próximo

RESUMO

As sementes de abóbora (*Cucurbita moschata* D.) contêm elevado teor de óleo, sendo em sua maioria composto por ácidos graxos insaturados. O uso de espectroscopia de infravermelho próximo (NIR) auxilia na análise química das sementes de forma fácil e rápida. O objetivo do trabalho foi construir e validar modelos de calibração multivariada utilizando NIR na análise química de teor de óleo e perfil de ácidos graxos de sementes de abóbora, e utilizar esses modelos em amostras com valores desconhecidos. Para o desenvolvimento dos modelos de calibração e validação foram utilizadas 324 amostras de sementes moídas de abóbora do programa de melhoramento genético da Universidade Federal de Viçosa. Parte das sementes foram analisadas pelos métodos convencionais de extração de óleo e cromatografia gasosa, enquanto que a outra parte foi moída para aquisição dos espectros NIR. Os modelos desenvolvidos foram utilizados na predição de 315 amostras de sementes de abóbora desconhecidas. Os modelos de calibração foram construídos utilizando regressão por quadrados mínimos parciais associado a seleção de variáveis de preditores ordenados (PLS-OPS). Os modelos apresentaram ajustes satisfatórios (valores R). Houve um ajuste linear entre os valores de referência e predição para todas as variáveis estudadas o que torna possível sua utilização na predição de amostras desconhecidas. O uso do NIR associado a métodos de calibração multivariada (PLS-OPS) possibilitou o desenvolvimento de um método rápido para análise do teor de óleo e perfil dos ácidos graxos de sementes moídas de abóbora, sendo capaz de fornecer resultados confiáveis.

Palavras-chaves: *Cucurbita moschata*, NIR, PLS-OPS, predição

1. INTRODUÇÃO

A abóbora (*Cucurbita moschata* Duchesne) é uma importante hortaliça do gênero *Cucurbita* devido à sua relevância socioeconômico e alimentar, e também ao alto valor nutricional de seus frutos e sementes. A espécie é fonte de vitaminas, sais minerais, polissacarídeos, carotenoides entre outros compostos benéficos à saúde (Chen e Huang, 2019).

As sementes da abóbora apresentam alto teor de óleo, aproximadamente 40%, dos quais, cerca de 70%, correspondem aos ácidos graxos insaturados, principalmente os ácidos linoleico e oleico (Veronezi e Jorge, 2015; Petkova e Antova, 2015). Além disso, o óleo possui potencial nutracêutico, sendo rico em compostos bioativos com propriedades funcionais (Bardaa et al., 2016; Cuco et al., 2019). Em razão dessas características, o óleo das sementes de abóbora é uma excelente alternativa a fontes lipídicas ricas em ácidos graxos saturados.

Os métodos de determinação do teor de óleo, como a extração por solvente, e análise do perfil dos ácidos graxos através da cromatografia gasosa são procedimentos que fornecem medições precisas. Entretanto, esses métodos apresentam desvantagens como alto custo, elevado consumo de tempo, necessidade de mão de obra especializada e utilização de solventes químicos tóxicos e geradores de resíduos (Sinelli et al., 2010). Deste modo, apesar da precisão, a utilização destas metodologias em programas de melhoramento genético torna-se indesejável.

Um método alternativo para predição do teor de óleo e perfil dos ácidos graxos é a espectrometria no infravermelho próximo (NIR). Esta técnica possui diversas vantagens como rapidez, fácil execução e a possibilidade de não destruição das amostras. Além disso, não requer a utilização de reagentes químicos, limita os erros de operação e utiliza amostra com reduzida preparação prévia. Adicionalmente, possibilita a análise simultânea de diversos parâmetros das amostras (Roque et al., 2017; Li et al., 2020). A metodologia NIR tem sido amplamente utilizada para determinar o teor de óleo e perfil de ácidos graxos em diversas culturas, incluindo soja (Jiang, 2020; Leite et al., 2020), milho (Fassio et al., 2015), girassol (Grunvald et al., 2014; Akkaya, 2018) e pinhão manso (Montes et al., 2013).

As informações espectrais NIR são muito sobrepostas o que dificulta sua interpretação a aplicação direta. Por isso, faz-se necessário a utilização de métodos quimiométricos, como a regressão multivariada. A regressão por quadrados mínimos parciais (PLS) é vastamente utilizada e tem se mostrado bastante efetiva na obtenção de modelos de predição utilizando o NIR.

Como o espectro NIR possui muitas variáveis, o uso de métodos de seleção de variáveis se torna importante para a seleção de regiões mais informativas e preditivas. O método de seleção de preditores ordenados (OPS), vem sendo utilizado juntamente com o método PLS,

por apresentar a capacidade de selecionar variáveis informativas em um conjunto de dados (Ferreira, 2015; Teófilo et al., 2009).

Diante do exposto, o objetivo deste estudo foi construir e validar modelos para predição do teor de óleo e o perfil de ácidos graxos de sementes de abóbora, utilizando espectroscopia NIR e regressão por quadrados mínimos parciais (PLS) utilizando o método de seleção de variáveis seleção de preditores ordenados (OPS). Os modelos de predição serão então utilizados na predição de amostras desconhecidas.

2. MATERIAL E MÉTODOS

2.1. Caracterização geral do experimento

Foram utilizadas 324 genótipos de populações de abóbora do programa de melhoramento genético de abóbora da Universidade Federal de Viçosa (UFV).

Os experimentos foram conduzidos no campo experimental do setor de Olericultura da Universidade Federal de Viçosa, em Viçosa – Minas Gerais, (20° 45' 14'' S, 42° 52' 53'' W). Todos os tratos culturais foram realizados de acordo com as recomendações para a cultura (Filgueira, 2008).

Ao atingir o ponto de colheita, três frutos de abóbora de cada genótipo foram selecionados ao acaso e suas sementes coletadas. As sementes foram lavadas em água corrente e secas em estufa de circulação de ar forçado à 40°C por 24 horas. Após a secagem, as sementes foram separadas em dois lotes, sendo um utilizado para o processo de extração de óleo e análise do perfil dos ácidos graxos, e o outro para o procedimento de leitura dos espectros no NIR.

2.2. Extração do óleo das sementes

O teor de óleo total das sementes foi determinado através da prensagem mecânica e via extrato etéreo, que quantificou o óleo remanescente da prensagem mecânica.

A prensagem mecânica foi realizada com uma prensa hidráulica com capacidade de 30 toneladas no Laboratório de Pós-Colheita da UFV. Aproximadamente 50 gramas de sementes de cada amostra ($MS_{inicial}$) foram secas em estufa de circulação de ar forçada por 24 horas à temperatura de 40°C. Posteriormente essas sementes foram prensadas até atingir a pressão de 20 toneladas e mantidas nesta pressão por 5 minutos.

A torta, produto obtido pela prensagem das sementes, foi pesada (M_{torta}) e o teor de óleo pela prensa mecânica ($TO_{inicial}$) foi determinado pela seguinte fórmula:

$$TO_{inicial} = \left(\frac{MS_{inicial} - M_{torta}}{MS_{inicial}} \right) * 100$$

Posteriormente essa torta foi triturada em moinho de facas tipo Willey com peneira de crivos de 1 mm, e submetida à extração do óleo residual pelo método do extrato etéreo com auxílio do extrator automático modelo ANKOM XT15. Cerca de 0,8 g da amostra moída foi acondicionada em saco de filtro modelo ANKOM XT4 com porosidade de 3 μ m. Os sacos de filtro foram devidamente numerados, lacrados e pesados (M_A). As amostras permaneceram em estufa à 105°C por 3 horas e foram novamente pesadas (M_B). Depois as amostras foram inseridas na cápsula de extração do equipamento, utilizando o éter de petróleo como solvente a uma temperatura de 90°C por 60 min. Ao final do processo de extração, as amostras foram transferidas para estufa a 105°C por 30 min para evaporação do éter e posterior pesagem (M_C).

O teor de óleo da torta (TO_{torta}) extraído através do método do extrato etéreo foi calculado pela fórmula:

$$TO_{torta} = \left(\frac{M_B - M_C}{M_A} \right) * 100$$

Já o teor de óleo total (%OL), que leva em consideração o teor de óleo extraído pela prensa e pelo extrato etéreo, foi obtido através da fórmula:

$$\%OL = \left[\frac{(MS_{inicial} * TO_{inicial}) + (M_{torta} * TO_{torta})}{MS_{inicial}} \right]$$

2.3. Análise de perfil dos ácidos graxos - Cromatografia gasosa

A análise do perfil de ácidos graxos do óleo das sementes de abóbora foi realizada por meio de cromatografia gasosa. Para isso, foi utilizado o óleo extraído via prensagem mecânica (método descrito acima) e a metodologia descrita por Bubeck et al., (1989) com adaptações.

Para a determinação do perfil de ácidos graxos foram usados tubos de tampa rosqueável onde foram adicionados 0,4mL de metóxido de sódio 1M a 5,0 μ L do óleo extraído, estes tubos foram tampados, posteriormente agitados no vortex e mantidos por 1 hora no banho maria a 30°C, depois foi adicionado 1ml de água milli Q e 1 ml de hexano. Os tubos foram novamente agitados e deixados em descanso por 1 hora, após este período coletou-se \pm 0,75ml da fase orgânica que foi transferida para outro tubo, onde foi adicionado sulfato de sódio anidro e posteriormente transferido a amostra para o tubo de injeção vial, para ser então realizada a análise cromatográfica.

Para a análise de cromatografia gasosa, foi usada o cromatógrafo a gás, modelo Shimadzu CG-17A, equipado com um detector por ionização de chama e uma coluna capilar

Carbowax (30 m × 0.25 mm). Foram inseridas amostras de 1,5 μL e como gás de arraste foi usado nitrogênio (15 psi). Durante as análises nos três primeiros minutos, a temperatura do forno foi mantida em torno de 60°C e em seguida elevada para 170°C em uma taxa de 5°C /min. Após nove minutos, a temperatura foi elevada para 250°C em uma taxa de 10°C /min. O injetor e o detector foram mantidos em temperaturas de 230 e 250°C, respectivamente. Os ácidos graxos foram identificados com base na comparação de seus tempos de retenção com o tempo de retenção de seus respectivos padrões. A quantificação individual dos ácidos graxos foi realizada calculando-se a porcentagem relativa do teor total de ácidos graxos, de acordo com o procedimento de normalização da área.

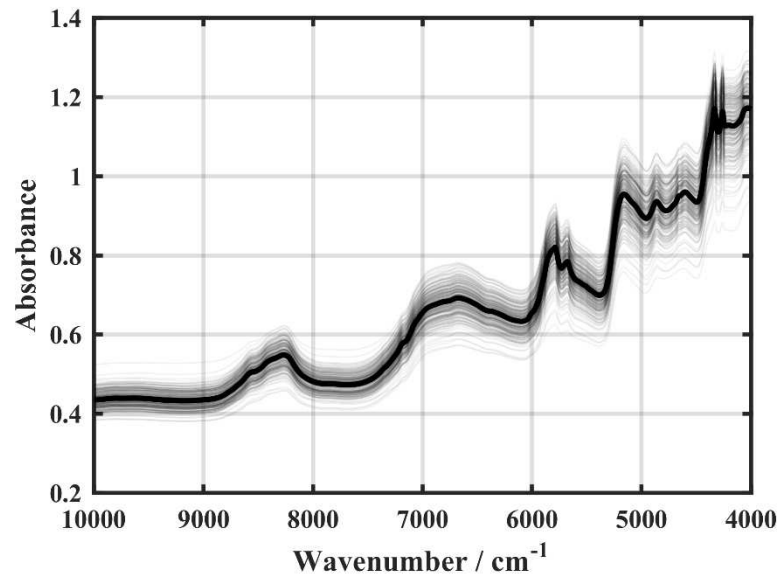
Foram identificados e quantificados os seguintes ácidos graxos: Palmítico (C16:0), Esteárico (C18:0), Oleico (C18:1), Linoleico (C18:2).

2.4. Análise espectral - NIR

As sementes do segundo lote foram moídas em moinho de facas tipo Willey com peneira de crivos de 1 mm. Após a moagem das sementes, os espectros NIR foram obtidos utilizando um instrumento Antaris II FT-NIR com transformada de Fourier e esfera de integração (Thermo Scientific). Os espectros foram obtidos utilizando software TQ Analysis. Os espectros NIR foram coletados a cada 30 seg com a faixa espectral de 10.000-4.000 cm^{-1} e resolução espectral de 3,8 cm^{-1} . Um espectro de referência de fundo foi adquirido a cada 15 amostras. Para cada amostra, 32 escaneamentos foram realizados e a média armazenada. Os espectros foram obtidos em módulo de reflectância difusa e os espectros armazenados como absorbância, $\log(1/R)$, onde R é a reflectância coletada.

Os espectros do NIR das amostras de sementes moídas de abóbora são apresentados na Figura 1.

Figura 1 – Espectros do NIR de sementes moídas de abóbora



2.5. Modelos de regressão multivariada

Os espectros foram importados para o Matlab 2019a (Math Works, Natick, EUA). Um modelo de regressão inversa ($y = Xb$) foi construído usando a regressão por quadrados mínimos parciais (PLS), onde X representa as variáveis independentes (espectros NIR), y as variáveis dependentes (valores de teor de óleo e ácidos graxos determinados) e b o coeficiente de regressão. Nas matrizes de variáveis independentes, cada linha corresponde a um espectro NIR de semente moída, e cada coluna corresponde a um número de onda, isto é, número de variáveis que serão usadas na modelagem.

De modo a encontrar o melhor modelo de previsão e eliminar as variações e os efeitos instrumentais, diversos pré-processamentos foram utilizados antes da construção do modelo. Além dos dados originais, foram estudados a centragem na média, alisamento, primeira derivada, segunda derivada, normalização e escalar pela variação normal padrão (Rinnan et al., 2009).

A regressão por quadrados mínimos parciais (PLS) foi utilizada para construir os modelos de previsão. A identificação e remoção dos outliers foi feita com base no resíduo padronizado de Student e nos valores de leverage (alavancagem). Após a remoção dos outliers, utilizando o algoritmo Kennard-Stone (1969), os dados foram separados em um conjunto de calibração, contendo 70% dos dados disponíveis, e em um conjunto de validação externa, contendo os 30% dos dados restantes.

A validação cruzada aleatória foi realizada com 10 divisões para selecionar o número de variáveis latentes (LV) e para otimizar os parâmetros dos modelos. Foram estudadas entre 1 e 10 variáveis latentes.

Após a obtenção dos melhores modelos, o método de seleção OPS (disponível em <http://www.deq.ufv.br/chemometrics>) foi utilizado para selecionar as variáveis mais importantes e que estão mais bem correlacionadas com o teor de óleo e o perfil dos ácidos graxos nas amostras de sementes moídas de abóbora. A seleção pelo método OPS é realizada ordenando as variáveis de acordo com vetores informativos e realizando uma busca sistemática pela melhor combinação de variáveis.

Todos os métodos OPS e vetores informativos disponíveis foram estudados utilizando uma janela inicial de 20 variáveis e um incremento de 10 variáveis. Informações mais detalhada sobre os métodos OPS podem ser encontradas na literatura (Teófilo et al., 2009; Roque et al., 2019).

Todos os modelos, assim como toda a otimização dos parâmetros, foram construídos utilizando apenas o conjunto de calibração, que após a definição dos parâmetros ótimos foi utilizado para prever o conjunto de validação externa. O conjunto de validação foi apenas utilizado após a obtenção do modelo para validar o mesmo e prever amostras desconhecidas

A qualidade dos modelos foi avaliada através da raiz quadrada do erro médio (RMSE) e pelo coeficiente de correlação (R) de acordo com as fórmulas:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{I_m} (y_i - \hat{y}_i)^2}{I_m}}$$

$$R = \frac{\sum_{i=1}^{I_m} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{I_m} (\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^{I_m} (y_i - \bar{y})^2}}$$

onde: y é o valor experimental, \hat{y} e $\bar{\hat{y}}$ são o escalar e o vetor dos valores estimados, respectivamente, e \bar{y} é um escalar dos valores médios de y e I_m é o número de amostras.

Quando é utilizada validação cruzada (CV), I_m é o número de amostras do conjunto de calibração (treinamento), o erro e o coeficiente de correlação são denotados RMSECV e Rcv, respectivamente. Quando I_m é o número de amostras de previsão, o erro e o coeficiente de correlação são chamados de RMSEP e Rp, respectivamente. Para verificação do ajuste do modelo foram usados como parâmetros os valores de RMSECV e Rcv. Já a capacidade de

predição do modelo foi verificada com base nos valores de RMSEP e Rp, assim como nos valores de %RE.

Foi calculado também a porcentagem do erro relativo (% RE), razão entre o erro absoluto e o valor medido, que é um parâmetro importante para verificar a precisão dos modelos construídos.

2.6. Predição de novos genótipos

Os modelos desenvolvidos utilizando NIR para o teor de óleo e perfil dos ácidos graxos de sementes de abóbora foram utilizados para a predição de 315 genótipos de abóbora (amostras desconhecidas). Para isso, foram coletados os espectros das sementes moídas desses genótipos, que tiveram seus valores estimados pelos modelos desenvolvidos.

3. RESULTADOS E DISCUSSÃO

3.1. Análise de referência do teor de óleo e composição de ácidos graxos

As médias, o desvio padrão e os valores mínimo e máximo para o teor de óleo e o perfil dos ácidos graxos das amostras analisadas de sementes moídas de abóbora estão apresentados na Tabela 1. As variáveis analisadas apresentaram ampla variação no intervalo dos dados analisados para as diferentes amostras estudadas, sendo assim, adequada para o desenvolvimento de calibrações NIR.

O teor de óleo apresentou média de 34,71%, com amplo intervalo de variação (Tabela 1). Gomes et al. (2020) em estudo com 95 genótipos de abóbora observaram maior variação no teor de óleo, porém com valor médio de apenas 18,52%, enquanto que Jarret et al. (2013) relataram teor médio de óleo de 28,7%, com máximo de 34,20%.

Com relação aos valores médios da composição dos ácidos graxos, as sementes de abóbora apresentaram maior teor médio para o ácido linoleico (48,75%) seguido pelo ácido oleico (26,89%), ácido palmítico (14,30%) e com menor média o ácido esteárico (10,04%) (Tabela 1).

Os resultados deste trabalho, considerando todos os ácidos graxos, se assemelham aos obtidos no estudo realizado com sementes de três variedades de *C. moschata* (Veronezi e Jorge, 2015). Boujema et al. (2020), relataram para *C. moschata* valores superiores de ácido linoleico (57,40%) e palmítico (17,39%), e inferiores para o ácido oleico (17,03%) e esteárico (7,26%). Por outro lado, Petkova e Antova (2015), obtiveram baixos valores para todos os ácidos graxos, exceto para o ácido palmítico (24,03%).

O alto teor de ácido linoleico pode promover a oxidação do óleo das sementes de abóbora. No entanto, este é um ácido graxo essencial, sendo nutricionalmente importante com diversos efeitos benéficos na saúde humana (Seymen et al., 2016; Reizeg et al., 2012). Por outro lado, a quantidade de ácido oleico presente no óleo proporciona alta estabilidade oxidativa, o que torna o óleo adequado para aplicações industriais (Stevenson et al., 2007).

Tabela 1 – Valores mínimos, máximos, médios e desvio padrão para o teor de óleo (%) e o perfil dos ácidos graxos (%) de sementes moídas de abóbora

Característica	Média	Mínimo	Máximo	Desvio Padrão
Teor de óleo	34.71	21.83	45.88	3.97
Esteárico	10.04	6.32	12.95	1.11
Linoleico	48.75	38.31	57.25	3.92
Oleico	26.89	16.46	38.17	4.34
Palmítico	14.30	12.14	16.41	0.79

3.2. Calibração e validação dos modelos NIR

A partir do desenvolvimento de modelos NIR para o teor de óleo e composição individual de ácidos graxos, os parâmetros de calibração, validação cruzada e validação externa são apresentados na Tabela 2. O número de variáveis latentes utilizadas para construir os modelos variou de 5 a 10.

Os valores mais baixos de erros RMSEP e RMSECV nos modelos PLS-OPS, para a maioria das variáveis, mostrou que o método OPS foi capaz de selecionar variáveis mais preditivas (Tabela 2).

Para todas as variáveis, o valor do erro relativo foi menor que 5%, o que mostra que os modelos são capazes de realizar a previsão do teor de óleo e perfil dos ácidos graxos por meio das sementes moídas com alta precisão, reduzindo significativamente o tempo gasto em análises e consumo de reagentes químicos.

Apesar dos valores de RMSEC e R_c serem próximo para a maioria das características estudadas, isto não exclui o uso de seleção de variáveis, pois com a seleção os modelos PLS-OPS são mais interpretativos, além de possuir a capacidade de identificar variáveis com mais capacidade preditiva.

Com relação a validação cruzada, os modelos PLS-OPS foram os que apresentaram maiores valores de R_{cv} , variando de 0,52 a 0,91, e menores valores de RMSECV, com exceção do ácido linoleico (Tabela 2). Valores altos de R_{cv} sugerem maior correlação com os valores da reta obtida, indicando maior ajuste entre os métodos laboratoriais e as análises em NIR.

Mesmo observando valor inferior de Rcv para o ácido oleico, os modelos de calibrações desenvolvidos obtiveram um bom poder preditivo. Os valores de Rp e RMSEP estão próximos aos de Rcv e RMSECV o que indica que os modelos não foram superestimados (Tabela 2).

O modelo PLS-OPS selecionou menos variáveis que o modelo completo, produzindo um modelo mais robusto e eficiente. Este modelo apresentou alta capacidade de predição ($R_p > 0,70$) para teor de óleo, ácido linoleico e oleico, com alta precisão em relação ao método de referência. Mesmo observando valores mais baixos de Rp para ácido esteárico e palmítico, os modelos ainda foram úteis na predição dessas características. Valores inferiores de Rp foram observados para ácido oleico (0,52) e linoleico (0,73) em estudo com sementes de pinhão manso, apesar disso, foi possível o desenvolvimento de modelo de predição para estas características (Vaknin et al., 2011).

Os valores mais baixos de Rc, Rcv e Rp (Tabela 2) obtidos para os ácidos palmítico e esteárico podem ser devidos a menor concentração destes ácidos nas sementes de abóbora, além do menor intervalo de variação e menor desvio padrão (Tabela 1). O mesmo foi observado por Baye e Becker (2004) em trabalho com sementes da planta oleaginosa *Vernonia galamensis*.

Tabela 2 – Parâmetros estatísticos de modelos completos (PLS) e parciais de mínimos quadrados ordenados (PLS-OPS) para espectros de espectroscopia no infravermelho próximo (NIR) para teor de óleo e perfil dos ácidos graxos

Caract.	Método	Var.	Calibração		Validação Cruzada		Validação Externa		LV
			RMSEC	Rc	RMSECV	Rcv	RMSEP	Rp	
Teor de óleo	PLS	1557	2.93	0.62	3.53	0.41	3.63	0.55	10
	PLS-OPS	290	2.61	0.72	3.14	0.57	3.16	0.70	10
Esteárico	PLS	1557	0.74	0.72	0.90	0.54	1.02	0.57	8
	PLS-OPS	980	0.77	0.69	0.88	0.56	1.04	0.55	8
Linoleico	PLS	1557	2.09	0.85	2.22	0.82	2.22	0.82	5
	PLS-OPS	510	2.12	0.84	2.27	0.82	2.20	0.83	5
Oleico	PLS	1557	1.70	0.92	2.11	0.88	2.34	0.85	10
	PLS-OPS	100	1.50	0.94	1.82	0.91	1.96	0.90	10
Palmítico	PLS	1557	0.62	0.62	0.68	0.51	0.66	0.55	6
	PLS-OPS	130	0.62	0.61	0.68	0.52	0.66	0.58	6

Caract.: característica; Var.: variáveis; LV: variáveis latentes; RMSEC: raiz quadrada média do erro de calibração; Rc: coeficiente de correlação de calibração; RMSECV: raiz quadrada média do erro da validação cruzada; Rcv: coeficiente de correlação de validação cruzada; RMSEP: raiz quadrada média do erro de predição; Rp: coeficiente de correlação de predição externa.

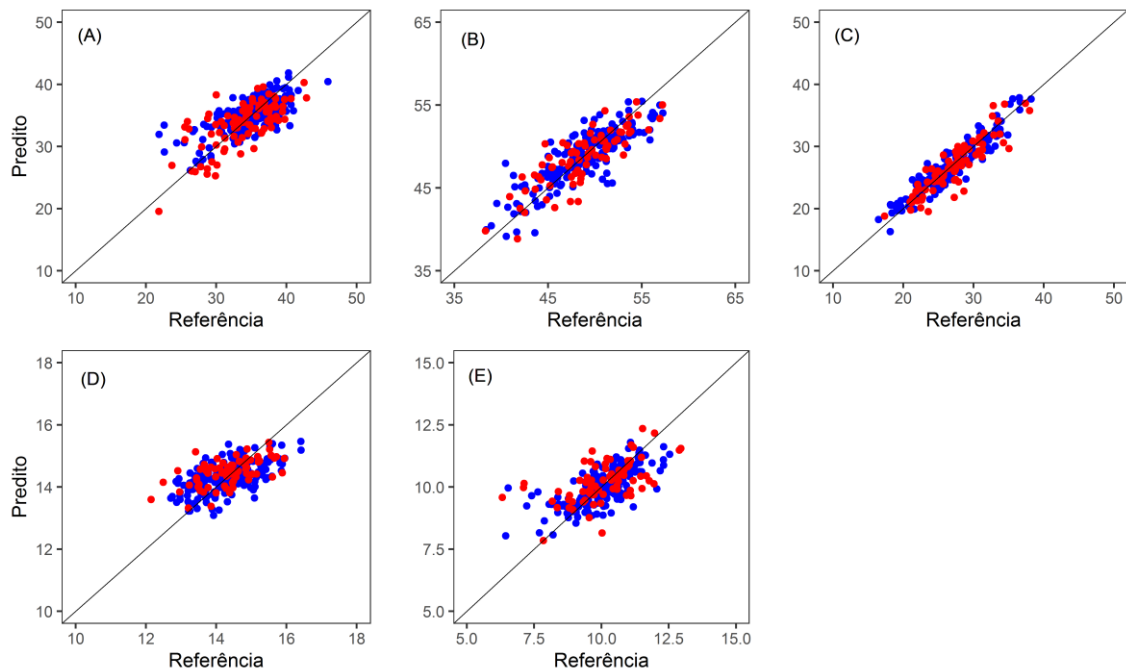
Uma comparação gráfica da relação entre os valores medidos e os valores preditos via NIR para teor de óleo e ácidos graxos é apresentada na Figura 2. Pode-se observar que há um

ajuste linear entre esses valores para todas as variáveis estudadas. Os modelos podem ser utilizados na discriminação entre amostras com diferentes teores de óleo e ácidos graxos, o que auxilia na tomada de decisão rápida no programa de melhoramento genético de plantas.

Quanto mais precisos os modelos preditivos, mais próximos os pontos se agrupam perto da correspondência teórica (linha sólida; correlação igual a 1) (Williams, 2001). O que demonstra que os valores previstos estão correlacionados aos valores reais.

De forma geral, os modelos construídos apresentaram uma boa correlação entre os espectros e os valores de referência, além de erros relativamente baixos.

Figura 2 – Valores de referência e valores preditos pela espectroscopia NIR para (A) Teor de óleo, (B) Ácido Linoleico, (C) Ácido Oleico, (D) Ácido Palmítico e (E) Ácido Esteárico. Pontos azuis representam os valores do conjunto de calibração e pontos vermelhos os valores do conjunto de validação externa



3.3. Predição do teor de óleo e ácidos graxos em amostras desconhecidas

Os modelos desenvolvidos utilizando NIR para as características estudadas foram utilizados para a predição de novos genótipos, cujas características eram desconhecidas. A representação gráfica dos valores preditos por espectroscopia NIR para teor de óleo e perfil dos ácidos graxos das sementes de 315 genótipos desconhecidos de abóbora são apresentadas na Figura 3.

A estimativa do teor de óleo por NIR agrupa a maioria dos genótipos (60,95%) com teores entre 32 e 40%, com destaque para dois genótipos com valores superiores a 40% de óleo

(Figura 3). O valor médio predito por espectroscopia para os 315 genótipos foi de 31,92% enquanto que as análises laboratoriais de 324 amostras foram de 34,71% (Tabela 1).

O valor médio de teor de óleo da semente de abóbora obtido neste trabalho foi superior ao encontrado em milho (7,94%), algodão (19,73%), soja (19,91%) e girassol (32,16%), e inferior somente ao óleo de canola, com 38,55% (Melchinger et al. 2017). Este resultado indica o potencial do óleo extraídos das sementes de abóbora frente a outras culturas oleaginosas.

Conforme as predições do NIR, 80,63% dos 315 genótipos apresentaram valor de ácido linoleico entre 40 e 52%. Além disso, 15 genótipos apresentaram teores acima de 52%. A média de ácido linoleico predita pelo NIR foi de 44,81% e pela análise de cromatografia gasosa foi de 48,75% (Tabela 1).

O teor estimado de ácido oleico está concentrado entre 24 e 32%, com 159 genótipos apresentando valores neste intervalo. Apesar da pouca quantidade, 17 genótipos de abóbora apresentaram teor de ácido oleico superior a 40%. A média predita por espectroscopia foi de 29,24%, por outro lado, as análises laboratoriais apresentaram 26,89% (Tabela 1).

A estimativa do ácido palmítico nos genótipos de abóbora apresentou menor variação (12,5-16%). Dos 315 genótipos analisados, 173 ficaram concentrados na faixa de 14 a 15%. A média predita por espectroscopia NIR e pela análise cromatográfica foram similares, sendo 14,38% e 14,30%, respectivamente.

Os valores estimados para o ácido esteárico variaram 8 a 17%. No entanto, aproximadamente 80% dos genótipos apresentaram valores entre 10 e 13%. Pela predição do NIR a média obtida foi de 11,71%, por outro lado, a análise de cromatografia gasosa apontou média de 10,04% (Tabela 1).

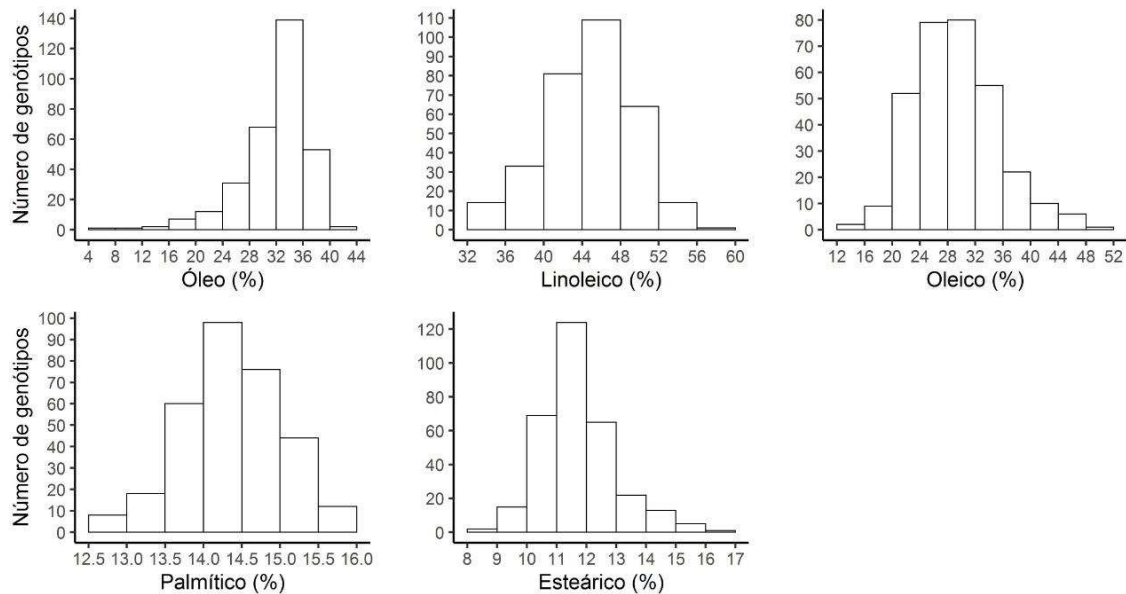
As médias estimadas neste estudo são similares aos obtidos por Bouazzaoui e Mulengi (2018) para ácido oleico (31,02%) e palmítico (15,01%) em sementes de abóbora. Em contrapartida, o ácido linoleico apresentou elevada média, com 50,75%, enquanto que o ácido esteárico foi observado valor médio de 3,22%.

Os valores médios dos ácidos oleico, palmítico e esteárico apresentados neste trabalho são superiores aos observados para as culturas do milho, da soja e do girassol. Por outro lado, com relação ao ácido linoleico, as sementes de abóbora possuem média inferior as todas culturas relatadas acima (Hoekman et al., 2012).

Segundo Seymen et al. (2016), o óleo extraído das sementes de abóbora apresenta potencial para utilização como óleo comestível, devido principalmente ao bom perfil de ácidos graxos e a elevada quantidade produzida quando comparado a outros óleos vegetais.

Adicionalmente, as sementes de abóbora podem ser utilizadas na biofortificação e tecnologia de alimentos em virtude dos compostos bioativos existentes (Veronezi, 2012).

Figura 3 – Distribuição dos valores preditos por espectroscopia NIR para teor de óleo e perfil dos ácidos graxos para genótipos de abóbora



4. CONCLUSÕES

O uso da espectroscopia no infravermelho próximo associado a métodos de calibração multivariada (PLS) possibilitou o desenvolvimento de um método rápido para análise das propriedades (teor de óleo e perfil dos ácidos graxos) de sementes moídas de abóbora. Os resultados mostram que os modelos desenvolvidos foram capazes de fornecer resultados confiáveis e que possibilitam a sua utilização nas etapas de seleção do programa de melhoramento de abóbora.

Os modelos desenvolvidos pelo NIR foram eficientes na predição de 315 genótipos desconhecidos de abóbora do programa de melhoramento da UFV, e poderão ser utilizados em outras populações de abóbora.

5. REFERÊNCIAS

AKKAYA, M. R. Prediction of fatty acid composition of sunflower seeds by near-infrared reflectance spectroscopy. **Journal of food science and technology**, v. 55, n. 6, p. 2318-2325, 2018.

BARDAA, S.; HALIMA, N. B.; ALOUI, F.; MANSOUR, R. B.; JABEUR, H.; BOUAZIZ, M.; SAHNOUN, Z. Oil from pumpkin (*Cucurbita pepo* L.) seeds: evaluation of its functional properties on wound healing in rats. **Lipids in health and disease**, v.15, n.73, p. 1-12, 2016.

BAYE, T; BECKER, H. C. Analyzing seed weight, fatty acid composition, oil, and protein contents in *Vernonia galamensis* germplasm by near-infrared reflectance spectroscopy. **Journal of the American Oil Chemists' Society**, v. 81, n. 7, p. 641-645, 2004.

BOUAZZAOUI, N.; MULENGI, J. K. Fatty acids and mineral composition of melon (*Cucumis Melo*) and pumpkin (*Cucurbita moschata*) seeds. *Journal of Herbs, Spices & Medicinal Plants*, v. 24, n. 4, p. 315-322, 2018.

BOUJEMAA, I.; BERNOUSSI, S. E. L.; HARHAR, H.; TABYAOUI, M. The influence of the species on the quality, chemical composition and antioxidant activity of pumpkin seed oil. **OCL**, v. 27, n. 40, p. 1 – 7, 2020.

BUBECK, D. M.; FEHR, W. R.; HAMMOND, E. G. Inheritance of palmitic and stearic acid mutants of soybeans. **Crop Science**, v.29, n. 3, p.652-656, 1989.

CHEN, L.; HUANG, G. Antioxidant activities of phosphorylated pumpkin polysaccharide. **International Journal of Biological Macromolecules**, v. 126, p. 743–746, 2019.

CUCO, R. P.; MASSA, T. B.; POSTAUE, N.; CARDOZO-FILHO, L.; DA SILVA, C. Oil extraction from structured bed of pumpkin seeds and peel using compressed propane as solvent. **The Journal of Supercritical Fluids**, 152, 104568, 2019.

FASSIO, A. S.; RESTAINO, E. A.; COZZOLINO, D. Determination of oil content in whole corn (*Zea mays* L.) seeds by means of near infrared reflectance spectroscopy. **Computers and Electronics in Agriculture**, v. 110, p. 171-175, 2015.

FERREIRA, M. M. C. Quimiometria – Conceitos, Métodos e Aplicações. Campinas, SP: Editora Unicamp, 2015. 493 f.

FILGUEIRA, F. A. R. Novo manual de olericultura: agrotecnologia moderna na produção e comercialização de hortaliças (3rd ed.). Viçosa, MG: Universidade Federal de Viçosa. 2008.

GOMES, R. S.; MACHADO JÚNIOR, R.; ALMEIDA, C. F.; CHAGAS, R. R.; OLIVEIRA, R. L.; DELAZARI, F. T.; DA SILVA, D. J. H. Brazilian germplasm of winter squash

(*Cucurbita moschata* D.) displays vast genetic variability, allowing identification of promising genotypes for agro-morphological traits. **PloS One**, v. 15, n. 6, e0230546, 2020.

GRUNVALD, A. K.; CARVALHO, C. G. P. D.; LEITE, R. S.; MANDARINO, J. M. G.; ANDRADE, C. A. D. B.; & SCAPIM, C. A. Predicting the oil contents in sunflower genotype seeds using near-infrared reflectance (NIR) spectroscopy. **Acta Scientiarum Agronomy**, v. 36, n. 2, p. 233-237, 2014.

HOEKMAN, S. K.; BROCH, A.; ROBBINS, C.; CENICEROS, E.; NATARAJAN, M. Review of biodiesel composition, properties, and specifications. **Renewable and sustainable energy reviews**, v. 16, n. 1, p. 143-169, 2012.

JARRET, R. L.; LEVY, I. J.; POTTER, T. L.; CERMAK, S. C.; MERRICK, L. C. Seed oil content and fatty acid composition in a genebank collection of *Cucurbita moschata* Duchesne and *C. argyrosperma* C. Huber. **Plant Genetic Resources: Characterization and Utilization**, v. 11, n. 2, p. 149–157, 2013.

JIANG, G-L. Comparison and Application of Non-Destructive NIR Evaluations of Seed Protein and Oil Content in Soybean Breeding. **Agronomy**, v. 10, n. 77, p. 1-15, 2020.

KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, v. 11, n. 1, p. 137-148, 1969.

LEITE, D. C.; CORRÊA, A. A. P.; JÚNIOR, L. C. C.; LIMA, K. M. G.; MORAIS, C. L. M.; VIANNA, V. F.; TEIXEIRA, G. H. A.; MAURO, A. O.; UNÊDA-TREVISOLI, S. H. Non-destructive genotypes classification and oil content prediction using near-infrared spectroscopy and chemometric tools in soybean breeding program. **Journal of Food Composition and Analysis**, v. 91, p. 103536, 2020.

LI, X.; ZHANG, L.; ZHANG, Y.; WANG, D.; WANG, X.; YU, L.; ZHANG, W.; LI, P. Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils. **Trends in Food Science & Technology**, v. 101, p. 172-181, 2020.

MATLAB (2019) MATLAB, Version 9.6.0 (R2019a) (Natick, Massachusetts: The MathWorks Inc).

MELCHINGER, A. E.; BÖHM, J.; UTZ, H. F.; MÜLLER, J.; MUNDER, S.; MAUCH, F. J. High-throughput precision phenotyping of the oil content of single seeds of various oilseed crops. **Crop Science**, v. 58, n. 2, p. 670-678, 2018.

MONTES, J. M.; TECHNOW, F.; BOHLINGER, B.; & BECKER, K. Grain quality determination by means of near infrared spectroscopy in *Jatropha curcas* L. **Industrial crops and products**, v. 43, p. 301-305, 2013.

PETKOVA, Zh Y.; ANTOVA, G. A. Changes in the composition of pumpkin seeds (*Cucurbita moschata*) during development and maturation. **Grasas y aceites**, v. 66, n. 1, p. 058, 2015.

REZIG, L.; CHOUAIBI, M.; MSAADA, K.; & HAMDY, S. Chemical composition and profile characterisation of pumpkin (*Cucurbita maxima*) seed oil. **Industrial Crops and Products**, v. 37, n. 1, p. 82-87, 2012.

RINNAN, A.; VAN DEN BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **Trends in Analytical Chemistry**, v. 28, n. 10, p. 1201-1222, 2009.

ROQUE, J. V.; CARDOSO, W.; PETERNELLI, L. A.; TEÓFILO, R. F. Comprehensive new approaches for variable selection using ordered predictors selection. **Analytica Chimica Acta**, v. 1075, p. 57-70, 2019.

ROQUE, J. V.; DIAS, L. A. S.; TEÓFILO, R. F. Multivariate calibration to determine phorbol esters in seeds of *Jatropha curcas* L. using near infrared and ultraviolet spectroscopies. **Journal of the Brazilian Chemical Society**, v. 28, n. 8, p. 1506-1516, 2017.

SEYMEN, M.; USLU, N.; TÜRKMEN, Ö.; AL JUHAIMI, F.; ÖZCAN, M. M. Chemical compositions and mineral contents of some hull-less pumpkin seed and oils. **Journal of the American Oil Chemists' Society**, v. 93, n. 8, p. 1095-1099, 2016.

SINELLI, N.; CASALE, M.; EGÍDIO, V.; OLIVERI, P.; BASSI, D.; TURA, D.; CASIRAGHI, E. Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy. **Food Research International**, v. 43, p. 2126 - 2131, 2010.

STEVENSON, D. G.; ELLER, F. J.; WANG, L.; JANE, J. L.; WANG, T.; INGLETT, G. E. Oil and tocopherol content and composition of pumpkin seed oil in 12 cultivars. **Journal of agricultural and food chemistry**, v. 55, n. 10, p. 4005-4013, 2007.

TEÓFILO, R. F.; MARTINS, J. P. A.; FERREIRA, M. M. C. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. **Journal of Chemometrics**, v. 23, n. 1, p. 32-48, 2009.

VAKNIN, Y.; GHANIM, M.; SAMRA, S.; DVASH, L.; HENDELSMAN, E.; EISIKOWITCH, D.; SAMOCHA, Y. Predicting *Jatropha curcas* seed-oil content, oil composition and protein content using near-infrared spectroscopy – A quick and non-destructive method. **Industrial Crops and Products**, v. 34, n. 1, p. 1029-1034, 2011.

VERONEZI, C. M. & JORGE, N. Chemical characterization of the lipid fractions of pumpkin seeds. **Nutrition & Food Science**, 45, 164-173, 2015.

VERONEZI, C. M.; JORGE, N. Bioactive compounds in lipid fractions of pumpkin (*Cucurbita* sp) seeds for use in food. **Journal of food science**, v. 77, n. 6, p. C653-C657, 2012.

WILLIAMS, P. C. (2001). Implementation of Near-infrared Technology. In: Williams, P. C.; Norris, K. H. (Eds.). Near Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemist, St. Paul, Minnesota, USA, p. 145-169.

CONSIDERAÇÕES FINAIS

Com o uso do índice de seleção FAI-BLUP foi possível selecionar genótipos superiores de abóbora com menor porte e maior produtividade de sementes visando a produção de óleo.

O NIR foi uma ferramenta eficaz na predição dos teores de óleo e perfil dos ácidos graxos de sementes moídas de abóbora. Fazendo com que o processo de análise fosse mais rápido e barato, além de não ser necessário a utilização de reagentes químicos.

Os modelos criados serão de suma importância em programas de melhoramento genético de abóbora com a finalidade de selecionar genótipos com maior teor de óleo e melhor perfil de ácidos graxos.