Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 261 (2021) 120036

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

FTIR spectroscopy with machine learning: A new approach to animal DNA polymorphism screening

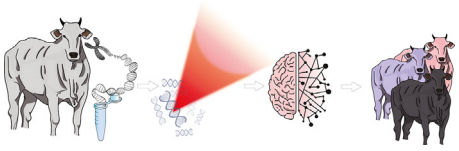
Thaynádia Gomes Rios [a], Gustavo Larios [b], Bruno Marangoni [b], Samuel L. Oliveira [b], Cícero Cena [b], Carlos Alberto do Nascimento Ramos [a,*]

[a] Faculdade de Medicina Veterinária e Zootecnia, Universidade Federal de Mato Grosso do Sul, 79070-900 Campo Grande, MS, Brazil
[b] Grupo de Óptica e Fotônica, Instituto de Física, Universidade Federal de Mato Grosso do Sul, 79070-900 Campo Grande, MS, Brazil

HIGHLIGHTS

- The FTIR associated to machine learning allowed for the identification of different genotypes;
- This approach can identify polymorphisms in animal DNA in a practical and rapid manner;
- The technique can be used as a screening methodology, reducing the costs of DNA sequencing by up to 90%.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 15 October 2020
Received in revised form 26 May 2021
Accepted 28 May 2021
Available online 1 June 2021

Keywords:
Genetic markers
Prospection
FTIR Spectroscopy
Animal
Machine learning
Multivariate analysis

ABSTRACT

Technological advances in recent decades, especially in molecular genetics, have enabled the detection of genetic DNA markers associated with productive characteristics in animals. However, the prospection of polymorphisms based on DNA sequencing is still expensive for the reality of many food-producing regions around the world, such as Brazil, demanding more accessible prospecting methods. In the present study, the Fourier transform infrared spectroscopy (FTIR) and machine learning algorithms were used to identify single nucleotide polymorphism (SNP) in animal DNA. The fragments of bovine DNA with well-known polymorphisms were used as a model. The DNA fragments were produced and genotyped by PCR-RFLP and classified according to the genotype (homozygous or heterozygous). FTIR spectra of DNA fragments were analyzed by principal component analysis (PCA) and machine learning algorithms. The best results exhibited 75–95% accuracy in the classification of bovine genotypes. Therefore, FTIR spectroscopy and multivariate analysis can be used as an alternative tool for prospecting polymorphisms in animal DNA. The method can contribute with studies to identify genetic markers associated with animal production and indirectly with food production itself, and reduce pressure on available natural resources.

1. Introduction

In the last decades, food production has grown significantly, whether of animal or vegetable origin. This increase occurred to attend to the population growth with the aid of technology in production systems [1,2]. Projections indicate that the world population will continue to grow, reaching almost 10 billion inhabitants in 2050 [3], which will further increase the demand for food. However, the intensification of agricultural production is likely to be mainly affected by the depletion of natural resources, such as water and land [4]. Thus, the use of technologies has been increasingly necessary for food production and should be based on the animal production paradigm with efficiency, quality, and productivity, in which genetic improvement programs prove to be effective [1].

With the development of molecular biology, new technologies have enabled the detection of polymorphisms or variations in

* Corresponding author.
E-mail address: carlos.nascimento@ufms.br (C. Alberto do Nascimento Ramos).

https://doi.org/10.1016/j.saa.2021.120036
1386-1425/

organisms' DNA sequences [5]. Single nucleotide polymorphisms (SNPs) have been associated with important productive characteristics in animals, such as a higher rate of daily weight gain [6] and greater tenderness of meat [7]. These specific polymorphisms are genetic markers and explored in breeding programs as a parent selection tool. Therefore, molecular genetics contributes to genetic improvement programs, increasing the prediction accuracy of genetic values of parent animals [8,9] and making animal production more efficient.

Nowadays, several molecular biology tools are available for the study of DNA polymorphisms. The Whole Genome Sequencing (WGS) studies based on NGS methods are undoubtedly the most efficient, considering that they enable the specific identification of polymorphisms in large DNA molecules. Although the costs of these analyses have undergone a significant reduction in recent years, they are still an obstacle when it is necessary to sequence the DNA of a large number of individuals. In turn, PCR-based techniques, such as PCR-RFLP, PCR-SSCP, and RAPD, present more accessible costs despite the herculean effort required to scan small DNA stretches used in prospecting studies.

Several studies of prospecting for genetic markers have been done to reduce costs and time, which list the genes most likely to be involved in the development of a phenotypic trait. The gene selection is based on the biochemical pathway of development of the trait. So, only a few genes along the genome are sequenced. However, even this method focused on one or a few genes is still costly and laborious due to the large size of the eukaryotic genes [10]. Thus, the development of a low-cost method or workflow that allows screening accurately which DNA stretches should be sequenced, as they are more likely to exhibit polymorphisms, could enable studies to prospect genetic markers in animals. FTIR spectroscopy has proven to be a versatile technique in qualitative and quantitative studies of several molecules [11,12]. Being simple, fast, and a non-destructive technique capable of analyzing small amounts of a sample [13], it has been used for different purposes, including chemical analysis, disease diagnosis, and identification of plant varieties [11,14,15].

Using genomic DNA as a target, the FTIR spectroscopy and multivariate analysis distinguished genetically modified varieties of rice and maize [16], as well as varieties of *Camellia reticulata* (ornamental plant) [15], and lineage of Chinese cabbage (*Brassica rapa* subsp. pekinensis) [17]. In the forensic field, FTIR also proved useful in the previous evaluation of bone samples for subsequent DNA examination and identification of victims [18]. In food inspection, the animal species were identified from DNA extracted from meat samples. DNA of bovine, sheep, fish, and swine were successfully distinguished using FTIR spectroscopy with least squares-discriminant analysis (PLS-DA) [19]. Thus, based on the versatility and practicality of the FTIR spectroscopy, the present study investigated the feasibility of this technique allied to machine learning as a new method for detecting polymorphism in animal DNA.

2. Experimental

2.1. Animal DNA samples and polymorphism

Genomic regions recognized for presenting single nucleotide polymorphisms (SNPs) were investigated to evaluate the proposed method. The region analyzed corresponds to an 891 bp bovine DNA fragment, located in intron 3 of the gene responsible for encoding the bovine growth hormone (bGH). In this region, a polymorphism caused by the C > T transition at position 1547 of the gene is known [20]. These polymorphisms, associated with characters of productive interest, were identified using PCR-RFLP as previously described [21]. A total of 60 bovine DNA samples were used, 20 from homozygous "CC" animals, 20 from heterozygous "CD" animals, and 20 from homozygous "DD" animals [21]. All animal DNA samples were obtained from the sample bank of the Molecular Biology Laboratory of FAMEZ/UFMS, Brazil.

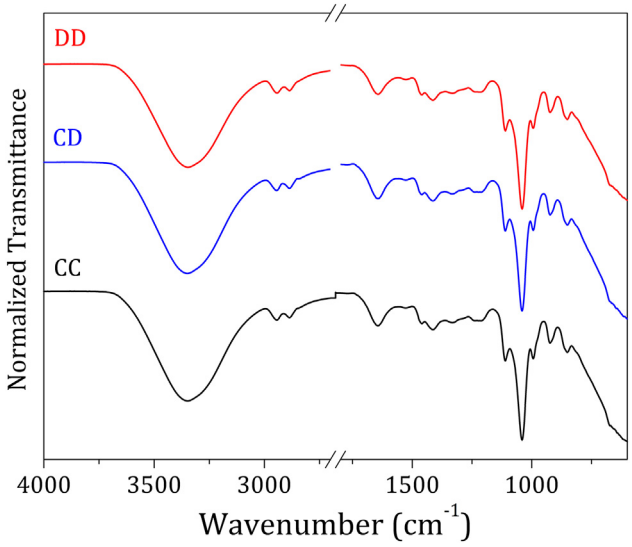


Fig. 1. Average FTIR spectra of bovine DNA samples. The legend refers to homozygous (CC, DD) and heterozygous (CD) alleles. The interval break from 1800 to 2700 cm^{-1} corresponds to a plateau, with no relevant information.

2.2. DNA sample preparation for FTIR spectroscopic measurements

The bovine DNA samples were amplified by PCR in a final volume of 100 μl, as described by [20]. The PCR products were visualized after electrophoresis on 1.5% agarose gel stained with GelRed (Biotium). A three-layer thick film was deposited on a silicon substrate (SiO_2) by casting method, followed by the drying at 40 °C for 30 min. Each layer was formed by 20 μl of bovine DNA sample deposition. The samples were measured in a Fourier transform infrared (FTIR) spectrophotometer (Spectrum 100, Perkin Elmer) using an attenuated total reflectance (ATR) accessory. The ATR-FTIR spectra were collected in the range from 4000 to 600 cm^{-1}, with a 4 cm^{-1} resolution and 10 scans. The average spectra were collected by measuring each film in 5 different spots.

2.3. Data analysis

The averaged FTIR spectra of the samples were investigated after standard normal variate (SNV) pretreatment. Principal Component Analysis (PCA) was applied to reduce the data dimensionality. The new variables are non-correlated among them and are called principal components (PC). From the first PCs plot in a cartesian graph (score plot graph), the data can reveal the clustering tendency of similar samples, enabling to check the predictive potential for sample classification. Additionally, by analyzing the weights (loadings) by which the original variable is multiplied to transform into the PCs, it is possible to examine the variables from the original data with higher variance in the spectral range. The PCA was used as an exploratory data analysis to evaluate the potential for sample differentiation.

After the PCA, supervised machine learning algorithms (ML) were applied to train the data set and generate a pattern recognition that allows the samples classification. The routines employed were the linear and quadratic discriminant analysis (DA); linear, quadratic, and cubic support vector machine (SVM); and fine, medium, cosine, cubic, and weighted K-nearest neighbor (KNN) [22].

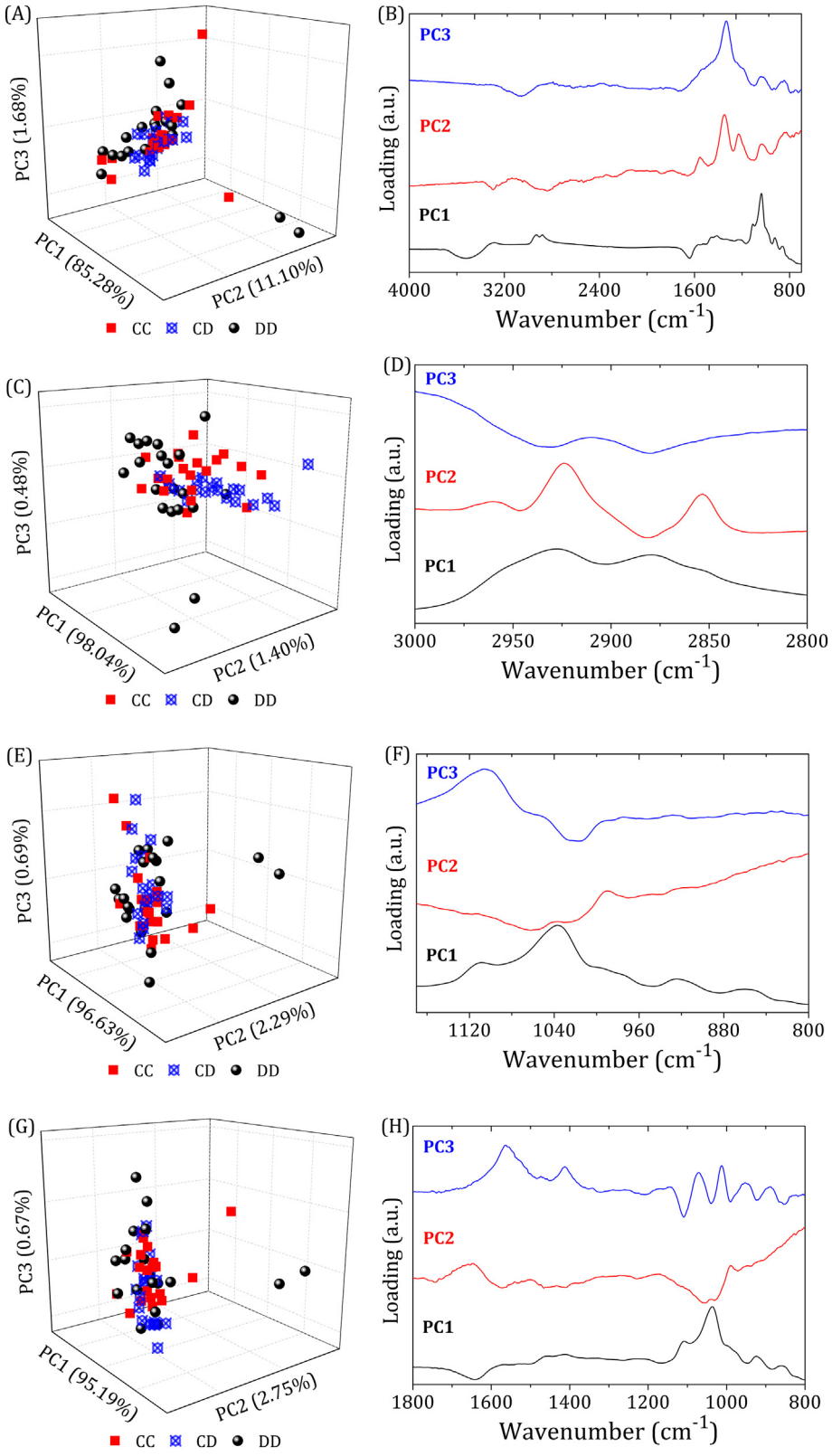


Fig. 2. Score plot graph (A, C, E, G) and loadings (B, D, F, H) from PCA of FTIR spectra of bovine DNA samples for different spectral ranges: (A, B) 4000–800 cm^{-1}, (C, D) 3000–2800 cm^{-1}, (E, F) 1170–800 cm^{-1}, and (G, H) 1800–800 cm^{-1}. The legend refers to homozygous (CC, DD) and heterozygous (CD) alleles.

The algorithms were performed with the aid of the Matlab2015b library. The ML was performed using the standard normal variate (SNV) spectra associated with reduced-dimensionality PCA, as described elsewhere [23,24].

The number of PCs used in the machine learning was chosen to avoid the over / underfitting regime. The leave-one-out cross-validation (LOO-CV) was performed to measure the accuracy of the classification model. The LOO-CV is a validation procedure in

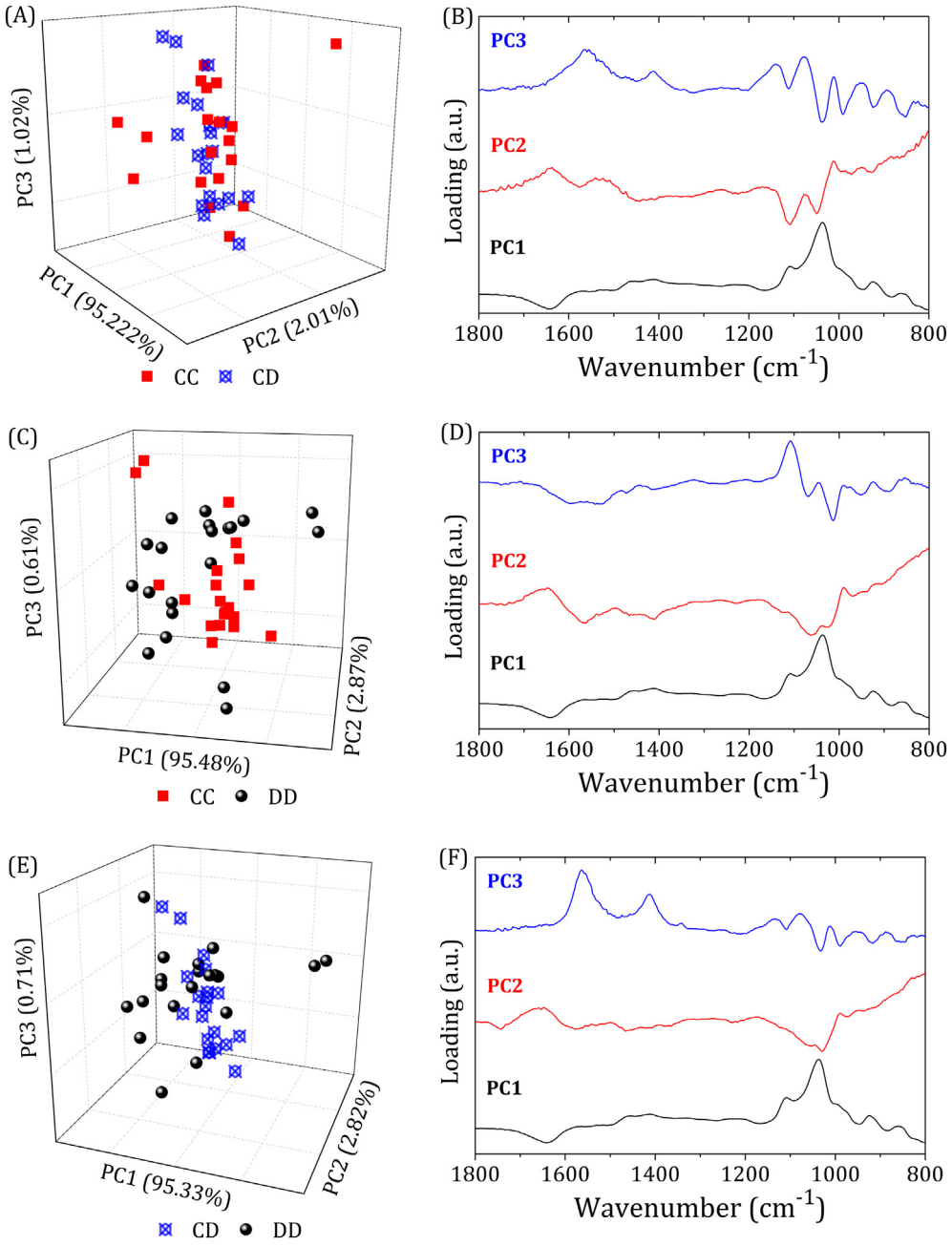


Fig. 3. Score plot graph (A, C, E) and loadings (B, D, F) from PCA of FTIR spectra of bovine DNA samples for the 1800–800 cm^{-1} range considering 1x1 combination groups. Combination tests were performed for (A, B) CCxCD, (C, D) CCxDD, and (E, F) CDxDD groups. The legend refers to homozygous (CC, DD) and heterozygous (CD) alleles.

which one sample is withdrawn from the training set, and the remaining one is used to train the model. The process is repeated until all samples are tested. The correct and incorrect classifications are recorded, returning the confusion matrix and the respective accuracy.

3. Results and discussion

3.1. Bovine DNA FTIR spectra

The bovine DNA samples exhibited a similar FTIR spectral signature (Fig. 1). The main vibrational bands were found in three different regions (800–1250 cm^{-1}; 1250–1500 cm^{-1}; and 1500–1800 cm^{-1}), which give important information on the molecular structure of DNA, providing details about nucleic acid interactions and conformations [25].

The first region (800–1250 cm^{-1}) is related to PO_4^{-2} group vibration and the deoxyribose stretching, highly sensitive to the nucleotide chain conformation [17,25,26]. The absorption band from 1250 to 1500 cm^{-1} is assigned to vibrational modes of aromatic rings in the nitrogenous base, such as pyrimidine (cytosine and thymine) and purine (adenine and guanine) [25,27,28]. The sugar wrinkling modes, glycosidic bonds rotation, and DNA conformation may affect their vibrational modes [25]. The third region from 1500 to 1800 cm^{-1} is attributed to the C = N, C = O, and C = C vibrational modes [25,29]. These modes are sensitive to the stacking and pairing of the molecules from the DNA nucleotides [25,30].

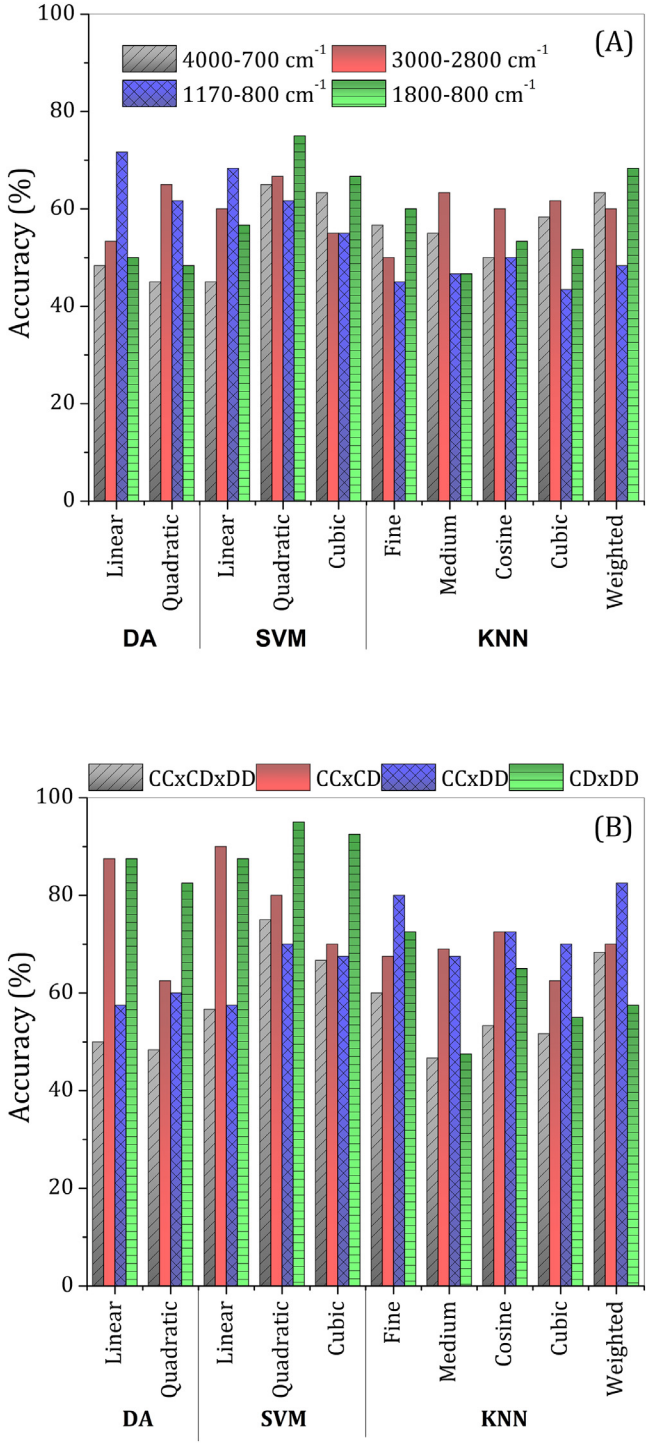


Fig. 4. Accuracy of the machine learning methods to distinguish three groups of bovine DNA samples (genotypes). (A) Machine learning tests performed to classify CCxDDxCD groups in different spectral ranges: (i) 4000–800 cm^{-1}, (ii) 3000–2800 cm^{-1}, (iii) 1170–800 cm^{-1}, and (iv) 1800–800 cm^{-1}, (B) Machine learning tests for the 1800–800 cm^{-1} range considering either 3 groups or 1x1 combination (CCxCD, CCxDD, and CDxDD). The legend refers to homozygous (CC, DD) and heterozygous (CD) alleles.

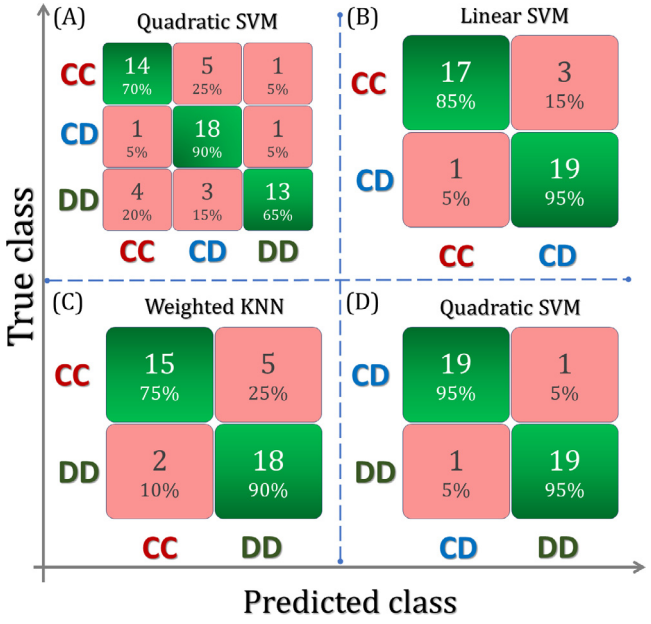


Fig. 5. Confusion matrix shows the best accuracy achieved using ML methods and different strategies to distinguish bovine DNA samples (genotypes). (A) Three groups tested simultaneously (quadratic SVM with 9 PCs); only 1x1 group of samples tested: (B) CCxCD (linear SVM with 11 PCs); (C) CCxDD (weighted KNN with 3 PCs); and (D) CDxDD (quadratic SVM with 22 PCs). All samples were tested by using standard normal variate (SNV) spectra associated with reduced-dimensionality PCA dataset. The legend refers to homozygous (CC, DD) and heterozygous (CD) alleles.

3.2. Principal component analysis

Fig. 2 shows the PCA results obtained by using average FTIR-SNV spectra for the three groups of bovine DNA samples in four different ranges: 4000–800 cm^{-1}, (ii) 3000–2800 cm^{-1}, (iii) 1170–800 cm^{-1}, and (iv) 1800–800 cm^{-1}. The score plots represent 98.06%, 99.92%, 99.61%, and 98.61% of the data variation for each spectral range, respectively (Fig. 2A, 2C, 2E, and 2G). No score plot showed a trend for clustering separation, regardless of the spectral ranges. The three groups of samples, two homozygous (CC and DD) and one heterozygous (CD), are clustered in only one formation, with a small dispersion along the PC3 axis.

The loading of PCs indicates that the first three ones contribute to the higher deviation of the data variance, mainly in the 1800–800 cm^{-1} range (Fig. 2B, 2D, 2F, 2H). Although the first three PCs explain more than 98% of the total variance for bovine DNA samples, the PCA analysis was not efficient to classify the samples and distinguish the three groups with different genotypes. The PCA can be successfully applied when there is some linear correlation between the features, but in this case, a high similarity between the mid-infrared spectral signature of the samples hindered the separation of the sample in the score plot.

As an alternative to improve clustering, groups in a 1x1 combination in the 1800–800 cm^{-1} range were tested. Fig. 3 displays the PCA results for CCxCD, CCxDD, and CDxDD group classification. The CCxCD score plot exhibits a high data correlation with no cluster separation or tendency (Fig. 3A). However, the CCxDD and CDxDD score plots exhibited higher score dispersion and a weak clustering trend (Fig. 3C and 3E). The loadings for all analyzed samples point out that the data variation is higher in the interval between 800 and 1250 cm^{-1}, region with a prominent absorption band assigned to the PO_4^{-2} stretching in nucleic acid molecules (Fig. 3B, 3D, 3F).

This high similarity can be explained considering that each sample was composed of a single type of DNA fragment or a mixture of two fragments differing from one or a few nucleotides. The bovine DNA region analyzed in the present study, for example, had 891 bp. Between different individuals, the only difference is a substitution of cytokine (C) per thymine (T) at a specific point [20]. This change represents about 0.1% (1/891) of the total DNA fragment and allows classifying individuals into different genotypes (homozygous or heterozygous). Animals with only fragments con-

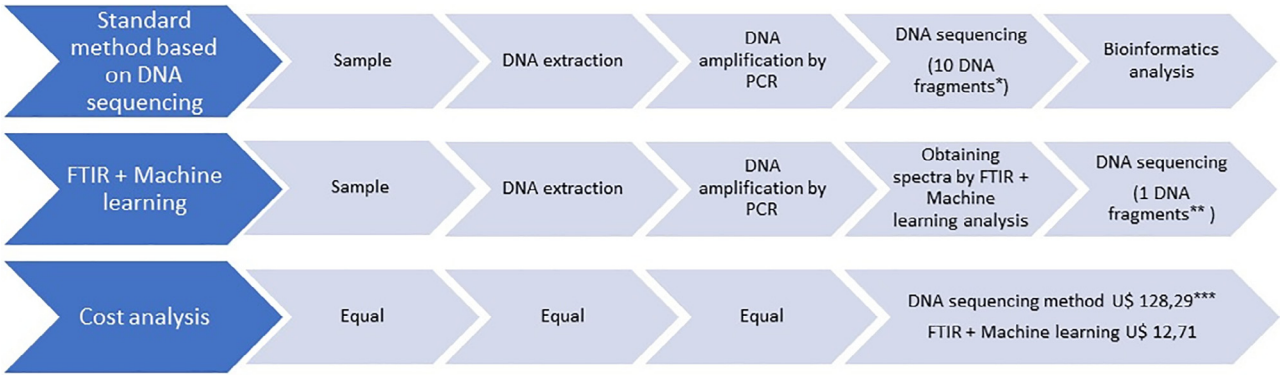


Fig. 6. Comparative schematic representation of the steps and respective costs for polymorphism detection in an animal gene of 10,000 bp using the traditional method or FTIR + machine learning approach. *Considering the limitation of the method for obtaining a DNA sequence of up to 1000 bp. **Considering only one polymorphism in a eukaryotic gene with 10,000 bp. ***Considering the average cost in Brazil of U$ 4.24 to sequence a fragment of up to 1000 bp using the Sanger method (US dollar quotation at R$ 5.19).

taining the nucleotide cytokine or thymine at position 1547 of the gene are classified as homozygous "CC or "DD". The bovine with fragments with cytokine and fragments with thymine in a 1/1 mixture is classified as heterozygous "CD" [21]. Therefore, the difference between the groups (genotypes) evaluated in the present study is small, so that it was not possible to distinguish them only with the aid of PCA of the FTIR spectra.

3.3. Machine learning tests

The machine learning (ML) algorithms were applied to the data after PCA processing in the normalized spectra to improve classification performance. Fig. 4A shows the overall accuracy from LOO-CV test results for three sample groups classification (CC, CD, and DD) considering four different spectral ranges. The highest overall accuracy of 75% was achieved, with 99.93% of data variance, using the first 9 PCs and quadratic SVM in the 1800–800 cm^{-1} range. Based on these results, ML training was performed by adopting a 1x1 group combination and the same 1800–800 cm^{-1} range as a strategy to improve the overall accuracy.

Fig. 4B shows the ML algorithms' results by testing the 1x1 group combination (CCxCD, CCxDD, and CDxDD), in the 1800–800 cm^{-1} range. A higher LOO-CV overall accuracy of 95% was obtained in the CDxDD classification by employing the first 22 PCs and quadratic SVM (99.99% data variance). The tests involving CCxCD provided the second higher accuracy (90%) using the first 11 PCs and linear SVM (99.95% of data variance). In comparison, the CCxDD classification led to the 82.5% accuracy adopting the first 3 PCs and the weighted KNN model, with a data variance of 98.96%.

Fig. 5 exhibits the confusion matrix determined by the LOO-CV tests for the best classification results by employing different strategies and the 1800–800 cm^{-1} range. The quadratic SVM was used to simultaneously classify the CC, CD, and DD bovine groups (Fig. 5A). Although the overall accuracy of the group classification was 75% (Fig. 4B), the CD classification reached 90%. In turn, the 1x1 classification led to an individual accuracy of more than 75% with a maximum of 95% by using weighted KNN, and linear or quadratic SVM, respectively (Fig. 5B, C, and D). Therefore, the ML algorithms classify the different bovine genotypes with a high agreement with the PCR-RFLP results.

In a traditional DNA polymorphism screening based on the Sanger DNA sequencing method [31], it is necessary to sequence the same DNA fragment several times to ensure a high level of confidence in the nucleotide indicated at each point [32]. Also, the Sanger method has limitations regarding the size of the DNA fragments that can be sequenced by reaction (approximately 1000 bp). In this context, to find a single-base polymorphism (SNP) in a eukaryotic gene with a size of 10,000 bp, it would be necessary to sequence 10 different fragments, each with 1000 bp, covering the entire length of the gene. With the FTIR-based method and machine learning, the sequencing costs and working time should be significantly reduced, considering that, after identifying the region most likely to contain a polymorphism, only it (one fragment of DNA) would need to be sequenced for confirmation of polymorphism. This approach could reduce costs by around 90%, according to estimates highlighted in Fig. 6.

The emergence of polymorphisms in DNA, changing the original codon, is a common and random event. This polymorphism can be called a genetic marker for a specific characteristic, which could signalize a harmful condition, such as cancer, or beneficial, such as resistance to a pathogen. However, the simple presence of polymorphisms does not indicate the exitance of a genetic marker for any characteristic. Most polymorphisms observed along with DNA molecules, especially in eukaryotes, do not have any known biological importance. Many of them occur in non-coding regions or do not involve alteration of the originally encoded amino acid. Thus, the approach proposed in the present study allows screening polymorphisms and evaluate whether these polymorphisms have a significant association with any relevant character before subjecting the DNA fragment to sequencing.

This work reports the first approach to use FTIR, multivariate analysis, and machine learning to identify DNA polymorphism. Due to the broad applicability and potential of the FTIR technique associated with robust statistical and computer analysis methodologies, the application for detecting mutations in DNA can extrapolate animal genetic improvement. The use of this approach for diagnosing mutations that cause genetic diseases also seems possible.

4. Conclusions

Fourier transform infrared spectroscopy (FTIR), associated with the multivariate analysis and machine learning, was able to classify different animal genotypes concerning specific DNA polymorphisms, providing an accurate and low-cost tool for animal DNA polymorphism screening. The PCA data analysis was not efficient to classify the samples by clustering, showing a significant data correlation. On the other hand, machine learning algorithms (KNN and SVM models) provided excellent results for bovine DNA classification. The best overall accuracy with high data variance was achieved considering the 1800–800 cm^{-1} range so that DNA nucleotides with vibrational modes in this range could have contributed to sample classification. Finally, the bovine DNA clas-

sification was improved up to 20% by testing the sample groups in a 1x1 manner.

CRediT authorship contribution statement

Thaynádia Gomes Rios: Conceptualization, Methodology, Investigation, Writing - original draft. **Gustavo Larios:** Investigation, Data curation. **Bruno Marangoni:** Investigation, Writing - review & editing. **Samuel L. Oliveira:** Writing - review & editing. **Cícero Cena:** Conceptualization, Methodology, Investigation, Writing - original draft. **Carlos Alberto do Nascimento Ramos:** Conceptualization, Methodology, Investigation, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors are grateful to the Fundação Universidade Federal de Mato Grosso do Sul for their support in carrying out the research.

References

[1] S.C. Fahrenkrug, A. Blake, D.F. Carlson, T. Doran, A. Van Eenennaam, D. Faber, C. Galli, Q. Gao, P.B. Hackett, N. Li, E.A. Maga, W.M. Muir, J.D. Murray, D. Shi, R. Stotish, E. Sullivan, J.F. Taylor, M. Walton, M. Wheeler, B. Whitelaw, B.P. Glenn, Precision genetics for complex objectives in animal agriculture, J. Anim. Sci. 88 (2010) 2530–2539, https://doi.org/10.2527/jas.2010-2847.

[2] H.C.J. Godfray, J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, J. Pretty, S. Robinson, S.M. Thomas, C. Toulmin, Food security: The challenge of feeding 9 billion people, Science (80-.) 327 (2010) 812–818, https://doi.org/10.1126/science.1185383.

[3] United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Highlights (ST/ESA/SER.A/423). https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf (accessed on May 01 2020).

[4] J.A. Foley, R. DeFries, G.P. Asner, C. Barford, G. Bonan, S.R. Carpenter, F.S. Chapin, M.T. Coe, G.C. Daily, H.K. Gibbs, J.H. Helkowski, T. Holloway, E.A. Howard, C.J. Kucharik, C. Monfreda, J.A. Patz, I.C. Prentice, N. Ramankutty, P.K. Snyder, Global consequences of land use, Science (80-.) 309 (2005) 570–574, https://doi.org/10.1126/science.1111772.

[5] R. Hedige, H.A. Ansari, G.F. Stranzinger, Chromosome banding and gene localizations support extensive conservation of chromosome structure between cattle and sheep, Cytogenet. Genome Res. 57 (1991) 127–134, https://doi.org/10.1159/000133131.

[6] M.H.A. Santana, Y.T. Utsunomiya, H.H.R. Neves, R.C. Gomes, J.F. Garcia, H. Fukumasu, S.L. Silva, P.R. Leme, L.L. Coutinho, J.P. Eler, J.B.S. Ferraz, Genome-wide association study for feedlot average daily gain in Nellore cattle (*Bos indicus*), J. Anim. Breed. Genet. 131 (2014) 210–216, https://doi.org/10.1111/jbg.12084.

[7] A.F.B. Magalhães, G.M.F. de Camargo, G.A. Fernandes, D.G.M. Gordo, R.L. Tonussi, R.B. Costa, R. Espigolan, R.M. de O. Silva, T. Bresolin, W.B.F. de Andrade, L. Takada, F.L.B. Feitosa, F. Baldi, R. Carvalheiro, L.A.L. Chardulo, L.G. de Albuquerque, Genome-Wide Association Study of Meat Quality Traits in Nellore Cattle, PLoS One 11 (2016) e0157845, https://doi.org/10.1371/journal.pone.0157845.

[8] A.R. Caetano, Marcadores SNP: Conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro, Rev. Bras. Zootec. 38 (2009) 64–71, https://doi.org/10.1590/S1516-35982009001300008.

[9] H. Carrer, A.L. Barbosa, D.A. Ramiro, Biotecnologia na agricultura, Estud. Avancados. 24 (2010) 149–164, https://doi.org/10.1590/S0103-40142010000300010.

[10] Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. The Complexity of Eukaryotic Genomes

[11] N.M. Amiali, M.R. Mulvey, J. Sedman, M. Louie, A.E. Simor, A.A. Ismail, Rapid identification of coagulase-negative staphylococci by Fourier transform infrared spectroscopy, J. Microbiol. Methods. 68 (2007) 236–242, https://doi.org/10.1016/j.mimet.2006.08.010.

[12] R. Liu, W. Tang, Y. Kang, M. Si, Studies on best dose of X-ray for Hep-2 cells by using FTIR, UV-vis absorption spectroscopy and flow cytometry, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 73 (2009) 601–607, https://doi.org/10.1016/j.saa.2009.03.001.

[13] T.R. Kahn, K.K. Fong, B. Jordan, J.C. Lek, R. Levitan, P.S. Mitchell, C. Wood, M.E. Hatcher, An FTIR investigation of flanking sequence effects on the structure and flexibility of DNA binding sites, Biochemistry 48 (2009) 1315–1321, https://doi.org/10.1021/bi8015235.

[14] D.K. Jangir, G. Tyagi, R. Mehrotra, S. Kundu, Carboplatin interaction with calf-thymus DNA: A FTIR spectroscopic approach, J. Mol. Struct. 969 (2010) 126–129, https://doi.org/10.1016/j.molstruc.2010.01.052.

[15] L. Qiu, Z. Wang, P. Liu, R. Liu, C. Cai, S. Fan, Fourier Transform Infrared Spectroscopy of the DNA of the Chuxiong Population of Camellia reticulata Lindl. of China, Spectrosc. Lett. 48 (2015) 120–127, https://doi.org/10.1080/00387010.2013.858267.

[16] K. Emura, S. Yamanaka, H. Isoda, K.N. Watanabe, Estimation for Different Genotypes of Plants based on DNA Analysis using Near-infrared (NIR) and Fourier-transform Infrared (FT-IR) Spectroscopy, Breed. Sci. 56 (2006) 399–403, https://doi.org/10.1270/jsbbs.56.399.

[17] S.Y. Song, E.Y. Jie, M.S. Ahn, I.H. Lee, I.S. Nou, B.W. Min, S.W. Kim, Fourier transform infrared (FT-IR) spectroscopy of genomic DNA to discriminate F1 progenies from their paternal lineage of Chinese cabbage (Brassica rapa subsp. pekinensis), Mol. Breed. 33 (2014) 453–464, https://doi.org/10.1007/s11032-013-9963-4.

[18] J.D. Fredericks, P. Bennett, A. Williams, K.D. Rogers, FTIR spectroscopy: A new diagnostic tool to aid DNA analysis from heated bone, Forensic. Sci. Int. Genet. 6 (2012) 375–380, https://doi.org/10.1016/j.fsigen.2011.07.014.

[19] Y. Han, L. Jian, Y. Yao, X. Wang, L. Han, X. Liu, Insight into rapid DNA-specific identification of animal origin based on FTIR analysis: A case study, Molecules 23 (2018) 2842, https://doi.org/10.3390/molecules23112842.

[20] J. Yao, S.E. Aggrey, D. Zadworny, J.F. Hayes, U. Kühnlein, Sequence variations in the bovine growth hormone gene characterized by single-strand conformation polymorphism (SSCP) analysis and their association with milk production traits in Holsteins, Genetics 144 (1996) 1809–1816, /pmc/articles/PMC1207730/?report=abstract (accessed September 18, 2020).

[21] A. Gorbani, R.V. Torshizi, M. Bonyadi, C. Amirinia, A MspI PCR-RFLP within bovin growth hormone gene and its association with sperm quality traits in Iranian Holstein bulls, African J. Biotechnol. 8 (2009) 4811–4816, http://www.academicjournals.org/AJB (accessed September 18, 2020).

[22] A.E. Casaril, C.G. Santos, B.S. Marangoni, S.M. Lima, L.H.C. Andrade, W.S. Fernandes, J.O.M. Infran, N.O. Alves, M.D.G.L. Borges, C. Cena, A.G. Oliveira, Intraspecific differentiation of sandflies specimens by optical spectroscopy and multivariate analysis, J. Biophoto. (2021), https://doi.org/10.1002/jbio.202000412.

[23] G. Larios, G. Nicolodelli, M. Ribeiro, T. Canassa, A. Rodrigues dos Reis, S. Oliveira, C. Alves, B. Marangoni, C. Cena, Soybean seed vigor discrimination by infrared spectroscopy and machine learning algorithms, Anal. Methods 12 (2020) 4303–4309, https://doi.org/10.1039/d0ay01238f.

[24] G.S. Larios, G. Nicolodelli, G.S. Senesi, M.C.S. Ribeiro, A.A.P. Xavier, D.M.B.P. Milori, C.Z. Alves, B.S. Marangoni, C. Cena, Laser-Induced Breakdown Spectroscopy as a Powerful Tool for Distinguishing High- and Low-Vigor Soybean Seed Lots, Food Anal. Methods 13 (2020) 1691–1698, https://doi.org/10.1007/s12161-020-01790-8.

[25] M. Banyay, M. Sarkar, A. Gräslund, A library of IR bands of nucleic acids in solution, Biophys. Chem. 104 (2003) 477–488, https://doi.org/10.1016/S0301-4622(03)00035-8.

[26] E.R. Blout, H. Lenormant, Changes in the infrared spectra of solutions of deoxypentose nucleic acid in relation to its structure, BBA - Biochim. Biophys. Acta 17 (1955) 325–331, https://doi.org/10.1016/0006-3002(55)90379-6.

[27] M.L.S. Mello, B.C. Vidal, Changes in the Infrared Microspectroscopic Characteristics of DNA Caused by Cationic Elements Different Base Richness and Single-Stranded Form, PLoS One 7 (2012) e43169, https://doi.org/10.1371/journal.pone.0043169.

[28] E. Taillandier, J. Liquier, Infrared spectroscopy of DNA, Methods Enzymol. 211 (1992) 307–335, https://doi.org/10.1016/0076-6879(92)11018-E.

[29] S.H. Brewer, S.J. Anthireya, S.E. Lappi, D.L. Drapcho, S. Franzen, Detection of DNA hybridization on gold surfaces by polarization modulation infrared reflection absorption spectroscopy, Langmuir 18 (2002) 4460–4464, https://doi.org/10.1021/la025613z.

[30] R. Ahmad, M. Naoui, J.F. Neault, S. Diamantoglou, H.A. Tajmir-Riahi, An FTIR spectroscopic study of calf-thymus dna complexation with al(iii) and ga(iii) cations, J. Biomol. Struct. Dyn. 13 (1996) 795–802, https://doi.org/10.1080/07391102.1996.10508892.

[31] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U. S. A. 74 (1977) 5463–5467, https://doi.org/10.1073/pnas.74.12.5463.

[32] M.C. Wendl, R.K. Wilson, Aspects of coverage in medical DNA sequencing, BMC Bioinfo. 9 (2008) 239, https://doi.org/10.1186/1471-2105-9-239.