Received: 11 October 2020 | Revised: 23 December 2020 | Accepted: 24 December 2020

DOI: 10.1002/jbio.202000412

FULL ARTICLE

JOURNAL OF BIOPHOTONICS

Intraspecific differentiation of sandflies specimens by optical spectroscopy and multivariate analysis

Aline E. Casaril[1] | Carlos G. Santos[2] | Bruno S. Marangoni[2] | Sandro M. Lima[3] | Luis H. C. Andrade[3] | Wagner S. Fernandes[1] | Jucelei O. M. Infran[1] | Natália O. Alves[1] | Moacir D. G. L. Borges[1] | Cicero Cena[2*] | Alessandra G. Oliveira[1*]

[1]Programa de Pós-Graduação em Doenças Infecciosas e Parasitárias, Faculdade de Medicina, UFMS – Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil

[2]Grupo de Ótica e Fotônica, Instituto de Física, UFMS – Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil

[3]Grupo de Espectroscopia Óptica e Fototérmica-GEOF, Centro de Estudos em Recursos Naturais- CERNA, Universidade Estadual de Mato Grosso do Sul-UEMS, Dourados, Brazil

***Correspondence**
Cicero Cena, Grupo de Ótica e Fotônica, Instituto de Física, UFMS – Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil.
Email: cicero.cena@ufms.br

Alessandra G. Oliveira, Programa de Pós-Graduação em Doenças Infecciosas e Parasitárias, Faculdade de Medicina, UFMS – Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil.
Email: alessandra.oliveira@ufms.br

Funding information
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: 001

Abstract

Lutzomyia longipalpis and *Lutzomyia cruzi* are the main sandflies species involved in the transmission of *Leishmania infantum* protozoan in Brazil. The morphological characteristics can be used for species identification of males specimens, while females are indistinguishable. Although, sandflies identification is essential to understand vectorial capacity, and susceptibility to infectious agents or insecticides, there is a lack of new strategies for specimen identification. In this study, Fourier transform infrared photoacoustic spectroscopy combined with multivariate analysis identified intraspecific differences between *Lutzomyia* populations. Successfully group clustering was achieved by principal component analysis. The main differences observed can be related to the protein content of the specimens. A classification with 100% accuracy was obtained using machine learning approach, allowing the identification of sandflies specimens.

KEYWORDS

classification, infrared spectroscopy, *Lutzomyia* sp., machine learning, multivariate analysis

1 | INTRODUCTION

Visceral leishmaniasis (VL) is a cosmopolitan zoonosis whose agents are heteroxenic digenetic protozoan of the *Leishmania* genus, which affect domestic and wild animals, and eventually, the human [1]. It is estimated that 50 000 to 90 000 new cases of VL occur worldwide annually. According to the World Health Organization, in 2018, more than 95% of new cases reported occurred in just 10 countries: Brazil, China, Ethiopia, India, Iraq, Kenya, Nepal, Somalia, South Sudan and Sudan [2]. In the American continent, 59 769 new cases were reported between 2001 and 2017. Brazil has the highest prevalence of VL cases (96%) [3]. In the Neotropical region, this vector-borne disease is caused by *Leishmania* (*Leishmania*) *infantum*. The epidemiological chain of this

zoonosis involves the hosts and insect vectors. In Brazil, *Lutzomyia longipalpis* and *Lutzomyia cruzi* are the two species that transmit this protozoan during females blood feeding [4–7].

L. longipalpis' adaptability to the urban environment has been considered one of the main determinants for the VL urbanization process in America. Among the factors that favor this scenario is the eclectic blood-feeding habit, which feeds by wild and domestic animals, such as dogs, horses, cows, chickens, birds, goats and humans [8, 9]. Due to its epidemiological importance, *L. longipalpis* is the most studied species in the Americas, and widely distributed from Mexico to Argentina [10].

The taxonomic classification of *L. longipalpis* and *L. cruzi* has been discussed once males are morphologically similar and females indistinguishable [11]. Also, it is believed there is a species complex with distinct populations of *L. longipalpis* in Brazil, based on biochemical, morphological, behavioral and molecular studies [12–14]. It is hypothesized that different populations may be responsible for different epidemiological profiles of the disease since sandflies have habits, behaviors and infection capacity intrinsic to the species [12].

The advance of technology and the mixing of physics, chemistry and biology have provided the essential tools for the elucidation of - biosystematics, population dynamics and phylogenetic relationships. Some studies have demonstrated the use of spectroscopic techniques for analyzing different biological systems, such as Fourier transform infrared photoacoustic spectroscopy (FTIR–PAS) [15–17].

FTIR-PAS is a useful technique that allows distinguishing different types of chemical bonds in a molecule [18]. Some studies have used this approach for analyzing different biological systems such as bacteria and fungus [19], ants [16], wasps [20], fishes [17] and plants [21]. FTIR spectroscopy is efficient in discriminating species and populations of insects as the technique based on vibrations related to functional groups present in the biomolecules present in the exoskeleton, such as carbohydrates, proteins, cuticular lipids, DNA and RNA [22]. In this study, we analyze the intraspecific differences between *L. longipalpis* and *L. cruzi* complex populations from different Brazilian regions by using infrared photoacoustic spectroscopy combined with multivariate analysis.

2 | MATERIALS AND METHODS

2.1 | Study area and period

We obtained 126 sandflies males using light traps (Falcão modified). The specimens, captured between August 2014 and December 2016, from five different Brazilian biomes, Figure 1: Amazon Rainforest (Belém, PA); Caatinga (Fortaleza, CE); Cerrado (Palmas, TO); Atlantic Rainforest (Recife, PE, and Jequié, BA); and Pantanal (Corumbá, MS), were identified based on morphological characteristics of the genitalia, head and thorax, as described by Galati [23].

2.2 | Fourier transform infrared photoacoustic spectroscopy

We investigated 21 sandflies specimens from each region by FTIR-PAS spectra, using a Thermo Nicolet Nexus 670 spectrometer, with photoacoustic detection, in the range of 4000 to 600 cm^{-1}, with 128 scans, and resolution of 16 cm^{-1}. The photoacoustic cell was purged with helium gas before each measurement. The standardization was kept by measuring a new background spectrum from a piece of carbon black after each 100 minutes, used as a reference.

The 21 specimens were measured in triplicate, by using simultaneously 03 entire specimens of sandflies in the sample holder, to improve the signal/noise ratio. The average FTIR-PAS spectra were analyzed and the main absorption bands were identified and assigned to respective molecular contributions.

2.3 | Data analysis

Interspecific differences among each group were evaluated by the FTIR-PAS spectra after standard normal variate (SNV) [24], which were dimensionally reduced by principal component analysis (PCA). After that, the data were trained and validated by machine learning algorithms. The PCA is an orthogonal linear transformation of the data that eliminates correlated variables and transforms it into a new set of uncorrelated variables, called principal components (PCs). Thus, the number of PCs resulting from the transformation is always smaller than the original set. The weights used to convert the variables into the PCs are called "loadings" and indicate the variables of the original set contribute the most to the transformation.

Machine learning algorithms are supervised training routines that use the data (original variables or PCs) to recognize patterns that allow them to differentiate among classes. The routines were developed using the Matlab R2015b library (The Mathworks INC., Natick) [25]. In this study, the number of PCs used in the machine learning training was optimized to avoid overfitting, so when the number of PCs was increased in the training model

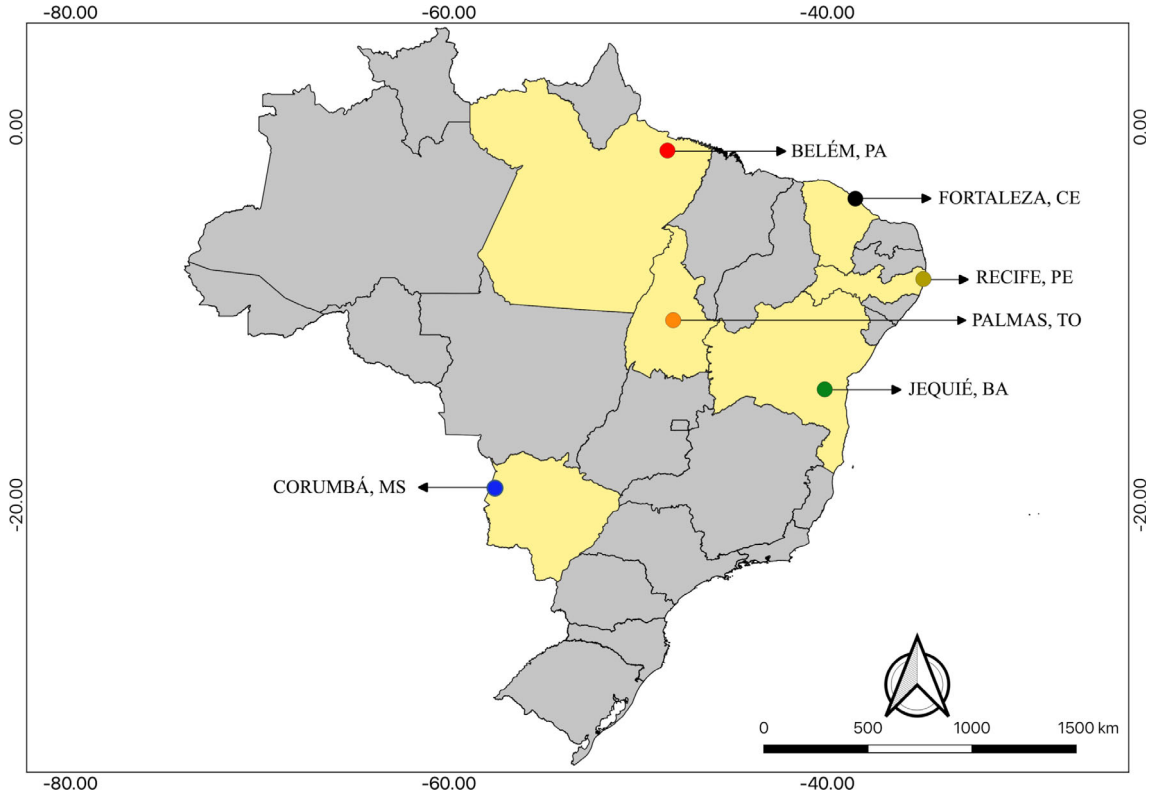

FIGURE 1 Geographic distribution of *Lutzomyia cruzi* (Corumbá, MS) and *Lutzomyia longipalpis* (Belém, PA; Fortaleza, CE; Recife, PE; Palmas, TO; Jequié, BA) specimens collected in Brazil. The PA, CE, PE, TO, BA and MS represent the Pará, Ceará, Pernambuco, Tocantins, Bahia e Mato Grosso do Sul, Brazilian federal units, respectively

the validation accuracy decreased (indicating overfitting). The same behavior was observed when the number of PCs was reduced (indicating underfitting).

The machine learning applied was based on different classifiers. The k-nearest neighbors (KNN) is a routine that arranges the variables in spatial distribution and classifies the sample according to the number of "k" nearest neighbors. The measured neighbor distance can be linear: fine KNN ($k = 1$) and medium KNN ($k = 10$), or nonlinear: cosine KNN ($k = 10$); cubic KNN ($k = 10$); and weighted KNN, where the closest neighbors have a higher influence in the classification. The support vector machine (SVM) organizes the data in spatial distribution and separates the classes by using hyperplanes, which can be shaped by different functions: linear, quadratic, and cubic. Discriminant analysis (DA) separates classes by molding boundaries among classes. Such delimiters can be linear (LDA) or quadratic (QDA). The ensemble classifiers (EC) combines several algorithms for classification. In the case of the EC-subspace discriminant and EC-subspace KNN, the routine randomly separates the set into subspaces and performs specific training for each subspace using the indicated classifiers (DA or KNN).

After the observation of PC´s loadings and the data trends, three main spectral regions associated with different molecular compositions were selected to be analyzed by machine learning algorithms. Finally, machine learning with PCA and SNV normalization spectra was applied and was optimized with the number of PCs. The validation process employed was the leave one out cross-validation (LOO-CV), where a sample is withdrawn from the training set and used to test the modeling predictive power. The process is repeated alternating the sample withdrawn to test all the sample sets.

3 | RESULTS AND DISCUSSION

Average FTIR-PAS spectra for sandflies from different Brazilian regions are shown in Figure 2. The analyzed samples exhibited similar FTIR-PAS absorption spectra in the mid-infrared range (4000-600 cm^{-1}), composed of vibrational contributions assigned to lipids, proteins, and carbohydrates molecules. In all samples, broadband is observed in the range 3700 until 3000 cm^{-1}, caused by H-bonded stretching from chitin, other polysaccharides and residual water [26].

The main bands, highlighted in Figure 2, were divided into four regions. Region I, from 3000 until 2800 cm^{-1} assigned to C—H stretching related methyl

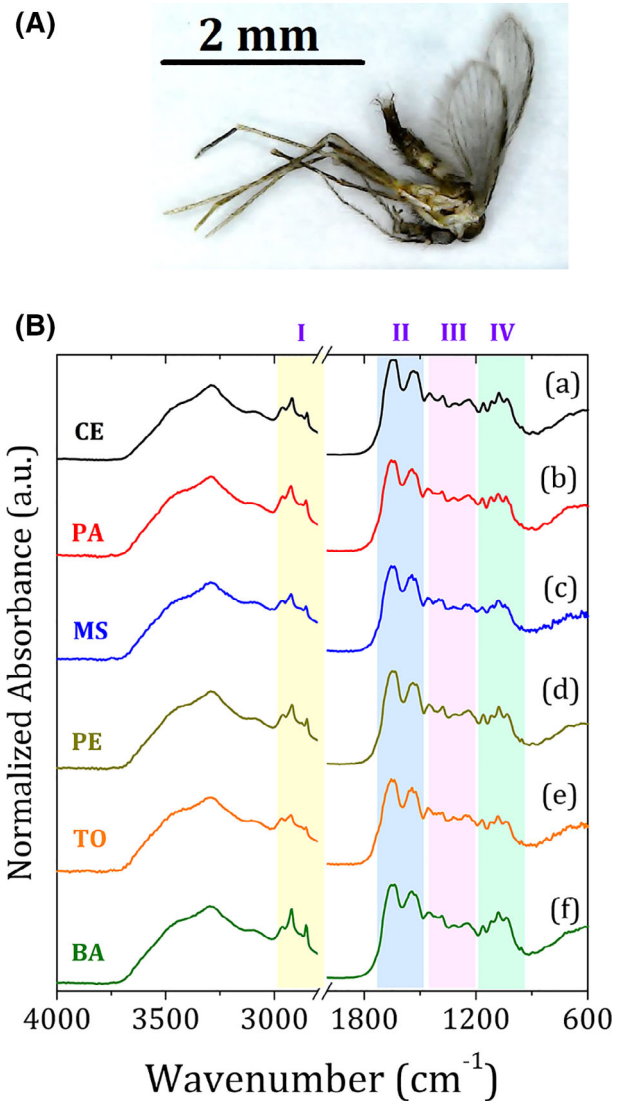

FIGURE 2 *Lutzomyia longipalpis* image and average Fourier transform infrared photoacoustic spectroscopy absorption spectra of sandflies from different regions of Brazil: (A) Fortaleza - Ceará, CE; (B) Belém - Pará, PA; (C) Corumbá - Mato Grosso do Sul, MS; (D) Recife - Pernambuco, PE; (E) Palmas - Tocantins, TO; (F) Jequié - Bahia, BA. The highlighted bands are associated with molecular signatures of lipids (I), proteins (II), lipids and proteins (III) and carbohydrates (IV)

groups and lipids. The two bands in region II correspond to amide 1 and amide 2 of proteins, located among 1600 and 1700 cm^{-1} and in the 1580 to 1510 cm^{-1} spectral range, respectively. The region III in the IR spectra ranging from 1480 cm^{-1} to 1340 cm^{-1} are methyl group bonds related to lipids and proteins. Besides, the phosphodiesterase groups (P═O) vibrate between 1300 and 1200 cm^{-1} and they are related to the sample nucleic acid. Finally, in region IV, the bands ranging from 1.100 to 1000 cm^{-1} are associated with functional carbohydrate groups [26, 27].

The dissimilar band shapes, Figure 2, can be assigned to the different molecular profiles of the specimens, probably promoted by environmental differences between their habitats. The specimens were collected from the Amazon rainforest, Atlantic rainforest, Caatinga and Savana-like Cerrado biomes. It is worth considering that the Atlantic and Amazon rainforest are similar, both are ombrophilous forests (broad and perennial vegetation and abundant and frequent rains), while the Cerrado is characterized as a biome of xeromorphic vegetation (with a seasonal climate and that covers aluminized leachate soils, consisting of low trees, widely spaced shrubs and grasses). These characteristics are probably responsible for altering the molecular constitution of insects and reflect on different molecular profiles [28–30].

3.1 | PCA analysis

Figure 3 shows the results of the PCA analysis of the FTIR-PAS data in three different ranges: (i) 4000 until 600 cm^{-1} (A and B), (ii) 3000 until 2800 cm^{-1} (C and D) and (iii) 2000 until 800 cm^{-1} (E and F). The left side of Figure 3A,C,E exhibits the score plot, and the right side (Figure 3B,D,F) the corresponding loading plot.

The score plot in Figure 3A,C, reveals a small tendency of clusterization. The score plot for the full range analysis, Figure 3A, represents 83.82% of the total data variance, it shows a formation of two different clusters, one associated with MS group (blue diamond), constituted by *L. cruzi*, and the other constituted by *L. longipalpis*. The scores associated with different groups of *L. longipalpis* seems to be forming a single cluster, mainly because the PA (red dots) and TO (orange dots) are dispersed, and each group is very close to each other. The corresponding loading, Figure 3B, shows that the main variance of the data is related to vibrational bands in the range (ii) associated with lipids molecules (region I in Figure 2), and in the range (iii) associated with proteins, carbohydrates and lipids molecules (regions II, III and IV in Figure 2), for all first three PCs.

The score plot considering a small region, from 3000 to 2800 cm^{-1}, represents 93.46% of the total data variance. Similar behavior, as described before, could be observed for clustering formation, as can be seen in Figure 3C, but here, a single cluster from BA group (green dots) identified distant from the other groups, that seems to form again one huge cluster with great dispersion and a mixture of the different dots. Analyzing the corresponding loadings, Figure 3D, the main deviation of the variance data for all first three PCs can be assigned to

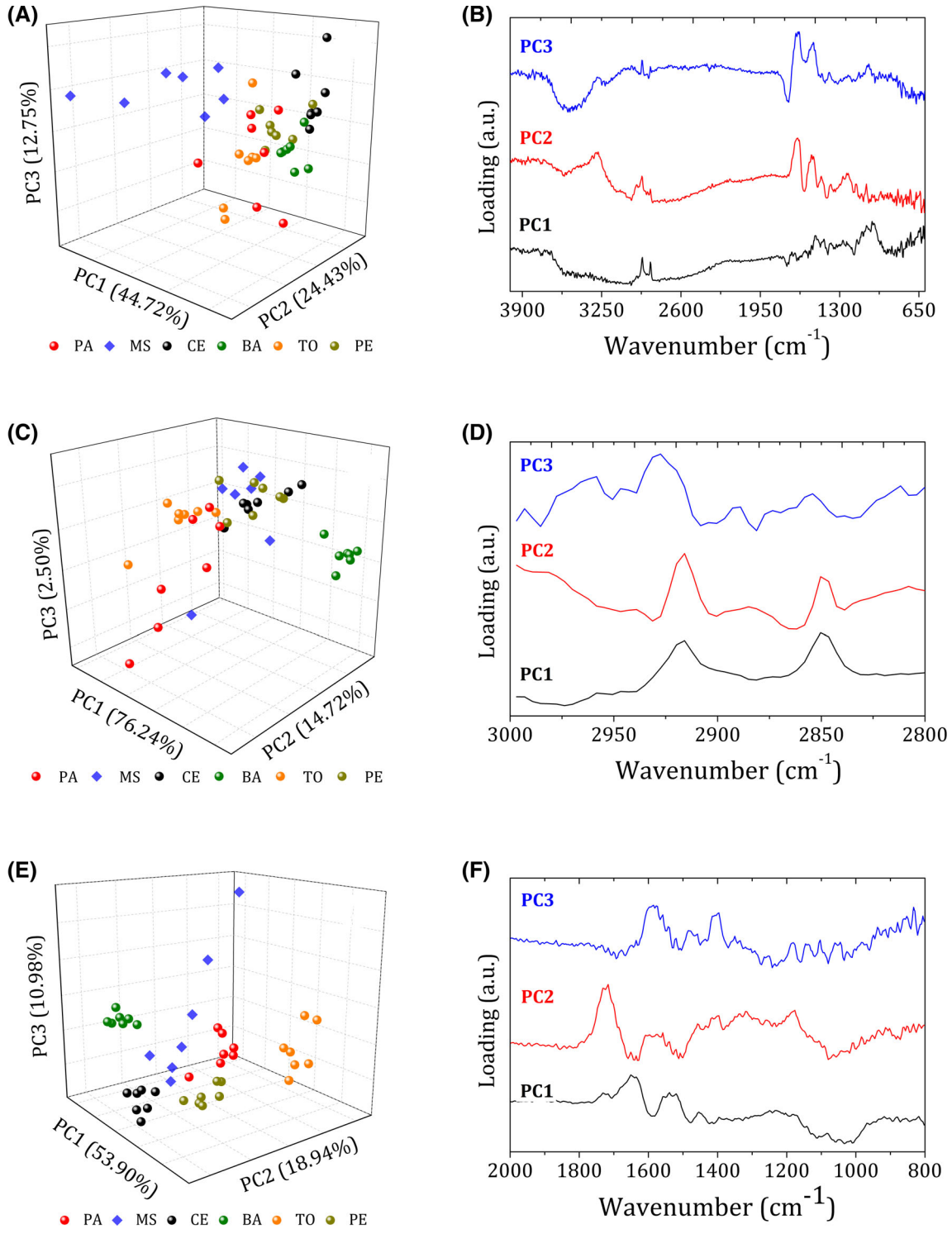

FIGURE 3 Score plot (A,C,E) and loadings (B,D,F) from principal component analysis (PCA) data analysis. (A,B) entire spectral range from 4000 to 600 cm^{-1}, (C,D) from 3000 to 2800 cm^{-1} and (E,F) from 2000 to 800 cm^{-1}

the C—H stretching vibrational bands, which is related to lipids molecules.

Successfully group clustering and separation were achieved by PCA in the range from 2000 until 800 cm^{-1}. The score plot, Figure 3E, presents a great cluster formation of *L. longipalpis*, while the cluster corresponding to *L. cruzi* (blue diamond) is quite dispersed. The loading, Figure 3F, shows that the most variance of the data correspond to vibrational bands between 1800 until 1500 cm^{-1}, these bands are assigned to proteins content, evidencing that the main differences between the groups can be related to its protein content of the specimens.

3.2 | Machine learning analysis

Figure 4 shows the accuracy achieved in the classification training by machine learning methods applied to the FTIR-PAS data, after SNV normalization and PCA processing. The process was optimized to return the best accuracy in classification and avoid under and overfitting [31]. In this way, the first seven PCs were selected to explain at around 95% of the total variance. The machine learning methods were performed in three different ranges: (a) 2000 until 800 cm^{-1} (green bars), (b) 4000 until 600 cm^{-1} (blue area) and (c) 3000 until 2800 cm^{-1} (red dot-line trace), in accordance to previously PCA analysis.

The highest classification accuracy (100%) were obtained in three different configurations, by analyzing: the full range (4000-600 cm^{-1}) applying the Ensemble Classifier algorithm with Subspace Discriminant; the selected range (2000-800 cm^{-1}) by applying the fine KNN and the cubic and quadratic SVM algorithms. The maximum accuracy achieved in the range from 3000 to 2800 cm^{-1} was 90.48% by using KNN and SVM algorithms. This result is probably related to the lack of information in the 2800 to 300 cm^{-1} range, where only lipid bands contributed to the spectra.

Although the full range training and validation provided 100% of accuracy, the result was achieved by using the ensemble classifier method which is a composition of several classifiers, requiring higher computational capacity [32]. On the other hand, KNN and SVM methods successfully classified the samples by using a reduced number of data, ranging from 2000 until 800 cm^{-1}. Following the PCA analysis, this range performs the best visual clusterization for the 3D score plot (Figure 3E) providing sufficient and important information to sample classification. The reduced set of variables, 2000 to 800 cm^{-1}, performs better than the full range, indicate that some variables, outside the selected range (2000-800 cm^{-1}), were not contributing significantly to the separation process. Such variables can present variance related to other characteristics instead of the differentiation of classes.

Figure 5 shows the confusion matrices obtained from the leave one out cross-validation tests by each method previously described. The classification performance of spectral regions from 3000 until 2800 cm^{-1}, Figure 5A, associated with lipids molecules, exhibited considerable confusion to classify samples from the MS region, corresponding to *L. cruzi*, a total of 42.9% of false-negative classification were obtained in this case. Additionally, just one sample was incorrectly classified in the TO groups, in which 85.7% of true positives were obtained.

On the other hand, successfully classification with 100% accuracy was obtained, Figure 5B, by analyzing the entire spectral range, 4000 until 600 cm^{-1}, by using the Ensemble Classifier method with Subspace discriminant. While the analysis of the selected range from 2000 until 800 cm^{-1} achieved 100% accuracy by using KNN and SVM methods.

Introgressive hybridization evidence between *L. longipalpis* and *L. cruzi* were observed in morphological studies [33–35], biochemical (sexual pheromones) [36], behavioral (*lovesongs*) [37] and molecular [11, 38], suggesting that the speciation process between these species is recent or is still occurring. Although these studies have provided important evolutionary information, our study, showed that FTIR-PAS combined with

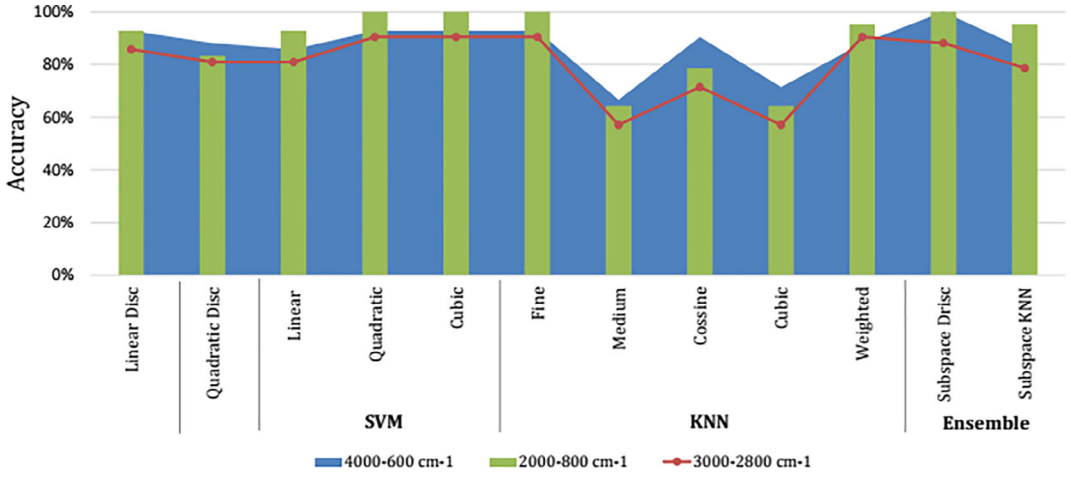

FIGURE 4 Accuracy of learning machine methods to discriminate and classify the sandflies groups. The green bars show the accuracy achieved in the range 2000 until 800 cm^{-1}, the blue area in the range 4000 until 600 cm^{-1}, and the red dot-line trace in the range 3000 until 2800 cm^{-1}

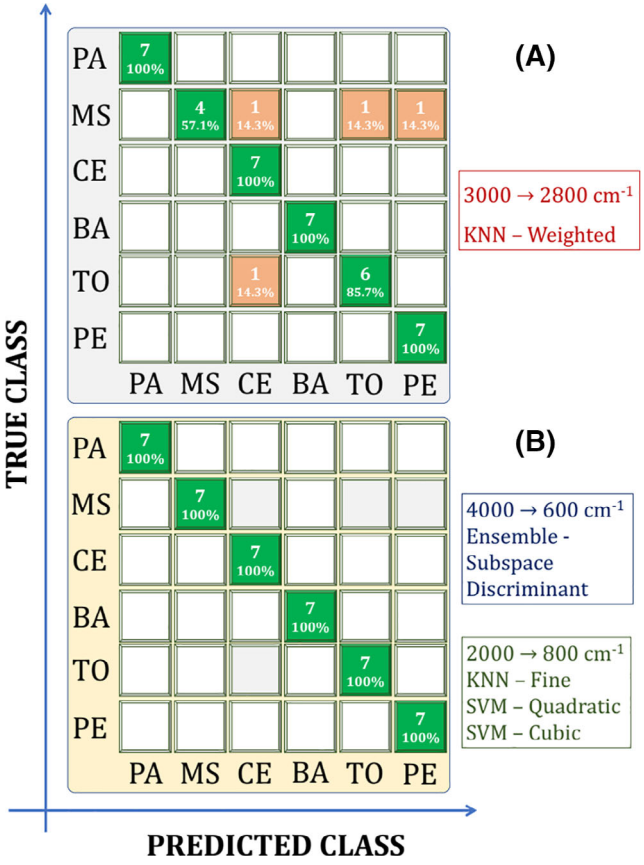

FIGURE 5 Confusion matrix for the performance of the best-supervised methods, by using seven principal components. (A) Selected range, 3000 until 2800 cm^{-1}, by using k-nearest neighbors (KNN) method with weighted function. (B) The entire range, 4000 until 600 cm^{-1}, by using the ensemble classifier method with subspace discriminant. Selected range, 2000 until 800 cm^{-1}, by using KNN with fine function, and support vector machine (SVM) method with quadratic and cubic functions

multivariate analyses can identify intraspecific differences between specimens. Similar results were found in studies involving male flies (Diptera: Sarcophagidae) at a specific level [22]. These results are due to the spectral characteristics derived from the biochemical composition of the cuticles of these insects, which may vary according to genetic and environmental factors [22, 39].

In morphological studies, the identification of sandflies consists of clarifying the insect, mounting it on a slide, and analyzing various characters that can be difficult to be visualized. It is common to lose part of the specimen during the clarification process, causing difficulties for species identification [40]. There is also the difficulty in visualizing structures by an inexperienced professional. Molecular studies, although they result in robust and reliable data, present disadvantages the high cost, the need for fresh material and require considerable effort and time [22].

On the other hand, FTIR-PAS combined with multivariate analysis proved to be an efficient alternative in the identification, since it presented high specificity and sensitivity both in terms of interspecific and population level. Another positive point for FTIR-PAS is that it does not require sample preparation, does not generate waste, and it is a fast and low-cost technique [22, 41].

4 | CONCLUSION

FTIR-PAS combined with multivariate analysis were able to identify intraspecific differences between sandflies specimens. FTIR-PAS associated with PCA can be used to identify sandflies specimens since the proper spectral range (2000-800 cm^{-1}) is selected. On the other side, FTIR-PAS combined with machine learning algorithm (ensemble classifier) successfully identified sandflies specimens by using the entire spectral range. These findings can be used to discriminate morphologically similar species of sandflies. This fact is relevant, as in areas where males are not captured, the identification of females cannot be performed, because they are indistinguishable. Besides, inexperienced people can often confuse the male of these species. FTIR-PAS proved to be a reliable and efficient methodology to assess differences in the chemical composition of sandflies and could be used as an auxiliary taxonomic tool, mainly to distinguish cryptic specimens.

ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001. The authors are grateful to Rafaella Albuquerque Silva (Ministério da Saúde), José Dilermando Andrade Filho (Instituto René Rachou FIOCRUZ Minas), Reginaldo Peçanha Brazil (FIOCRUZ), Ana Caroline Moura Rodrigues (Laboratório de Doenças Parasitárias, Universidade Estadual do Ceará) and Gustavo S. Larios.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Conceptualization: Aline E. Casaril, Cicero Cena and Alessandra G. Oliveira; Sandflies collection and screening: Aline E. Casaril, Wagner S. Fernandes, Jucelei O. M. Infran, Natália O. Alves, Moacir D. G. L. Borges; Data acquisition: Sandro M. Lima and Luis H. C. Andrade; Formal analysis: Carlos G. Santos, Bruno S. Marangoni, and Cicero Cena; Funding acquisition: Alessandra G. Oliveira; Investigation: Aline E. Casaril and Carlos

G. Santos; Methodology: Cicero Cena and Bruno S. Marangoni; Project administration: Alessandra G. Oliveira; Machine learning analysis: Bruno S. Marangoni; Visualization; Writing - original draft: Aline E. Casaril and Cicero Cena; Writing - review: Wagner S. Fernandes, Jucelei O. M. Infran, Natália O. Alves, Moacir D. G. L. Borges, Alessandra G. Oliveira, and Cicero Cena; Editing: Cicero Cena.

DATA AVAILABILITY STATEMENT

Data will be avaliable under request.

ETHICS STATEMENT

A permanent license for collecting and transporting zoological material (Protocol 25 592–1) was obtained on the behalf of Ph.D. Alessandra Gutierrez de Oliveira, issued by the System of Authorization and Information on Biodiversity of the Brazilian Institute of Environment and Renewable Natural Resources (Sisbio/IBAMA).

ORCID

Cicero Cena https://orcid.org/0000-0001-8766-6144

REFERENCES

[1] R. Lainson, E. F. Rangel, *Mem. Inst. Oswaldo Cruz* **2005**, *100*, 811.
[2] World Health Organization, Leishmaniasis, https://www.who.int/news-room/fact-sheets/detail/leishmaniasis, (accessed: July **2020**).
[3] Organización Panamericana de la Salud, Leishmanioses, *Informe Epidemiológico das Américas*, OPS, Washington, DC **2019**.
[4] L. M. Deane, M. P. Deane, *Hospital* **1954**, *47*, 75.
[5] M. P. Deane, L. M. Deane, *Hospital* **1954**, *45*, 697.
[6] E. F. de Oliveira, E. T. Oshiro, W. S. Fernandes, A. M. T. Ferreira, A. G. de Oliveira, E. A. B. Galati, *Am. J. Trop. Med. Hyg.* **2017**, *96*, 178.
[7] E. F. de Oliveira, E. T. Oshiro, W. S. de Fernandes, P. G. Murat, M. J. de Medeiros, A. I. Souza, A. G. de Oliveira, E. A. B. Galati, *PLoS Negl. Trop. Dis.* **2017**, *11*, e0005401.
[8] A. G. de Oliveira, A. M. Marassá, C. A. Consales, M. E. C. Dorval, C. E. Fernandes, G. R. de Oliveira, R. P. Brazil, E. A. B. Galati, *Acta Trop.* **2008**, *107*, 238.
[9] E. F. Rangel, M. L. Vilela, *Cad. Saude Publica* **2008**, *24*, 2948.
[10] A. T. Peterson, L. P. Campbell, D. A. Moo-Llanes, B. Travi, C. González, M. C. Ferro, G. E. M. Ferreira, S. P. Brandão-Filho, E. Cupolillo, J. Ramsey, A. M. C. Leffer, A. Pech-May, J. J. Shaw, *Int. J. Parasitol.* **2017**, *47*, 667.
[11] M. F. C. Santos, P. E. M. Ribolla, D. P. Alonso, J. D. Andrade-Filho, A. E. Casaril, A. M. T. Ferreira, C. E. S. Fernandes, R. P. Brazil, A. G. Oliveira, *PLoS One* **2013**, *8*, e74268.
[12] A. E. Casaril, D. P. Alonso, K. G. Franco, M. V. N. Alvarez, S. P. G. Barrios, W. S. de Fernandes, J. O. M. de Infran, A. C. M. Rodrigues, P. E. M. Ribolla, A. G. de Oliveira, *PLoS One* **2019**, *14*, e0223277.
[13] N. A. Souza, R. P. Brazil, A. S. Araki, *Mem. Inst. Oswaldo Cruz* **2017**, *112*, 161.
[14] C. N. Spiegel, D. B. d. S. Dias, A. S. Araki, J. G. C. Hamilton, R. P. Brazil, T. M. Jones, *Parasit. Vectors* **2016**, *9*, 580.
[15] W. F. Antonialli Junior, Y. R. Súarez, T. Izida, L. H. C. Andrade, S. M. Lima, *Genet. Mol. Res.* **2008**, *7*, 559.
[16] W. F. Antonialli Junior, S. M. Lima, L. H. C. Andrade, Y. R. Súarez, *Genet. Mol. Res.* **2007**, *8*, 492.
[17] F. S. de Almeida, C. A. Santana, D. M. V. Lima, L. H. C. Andrade, Y. R. Súarez, S. M. Lima, *Infrared Phys. Technol.* **2016**, *76*, 303.
[18] C. Santos, M. E. Fraga, Z. Kozakiewicz, N. Lima, *Res. Microbiol.* **2010**, *161*, 168.
[19] J. Irudayaraj, H. Yang, S. Sakhamuri, *J. Mol. Struct.* **2002**, *606*, 181.
[20] E. F. Neves, L. D. Lima, D. Sguarizi-Antonio, L. H. C. Andrade, S. M. Lima, S. E. Lima-Junior, W. F. Antonialli-Junior, *Neotrop. Entomol.* **2019**, *48*, 1030.
[21] M. S. Pontes, R. Grillo, D. E. Graciano, W. F. Falco, S. M. Lima, A. R. L. Caires, L. H. C. Andrade, E. F. Santiago, *Ecotoxicol. Environ. Saf.* **2019**, *180*, 526.
[22] T. M. Barbosa, L. A. S. de Lima, M. C. D. dos Santos, S. D. Vasconcelos, R. A. Gama, K. M. G. Lima, *Acta Trop.* **2018**, *185*, 1.
[23] E. A. B. Galati, in *Brazilian Sand Flies: Biology, Taxonomy, Medical Importance and Control* (Eds: E. F. Rangel, J. J. Shaw), Springer International Publishing, Cham **2018**, p. 9.
[24] Å. Rinnan, F. van den Berg, S. B. Engelsen, *TrAC Trends Anal. Chem.* **2009**, *28*, 1201.
[25] G. Larios, G. Nicolodelli, M. Ribeiro, T. Canassa, A. R. Reis, S. L. Oliveira, C. Z. Alves, B. Marangoni, C. Cena, *Anal. Methods* **2020**, *12*, 4303.
[26] I. Ferreira, J. Ferreira-Strixino, M. L. Castilho, C. B. L. Campos, C. Tellez, L. Raniero, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2016**, *152*, 397.
[27] M. A. Herrera-Valdez, J. Lega, *J. Theor. Biol.* **2011**, *270*, 164.
[28] Sistema Nacional de Informações Florestais, Bioma Cerrado, http://snif.florestal.gov.br/pt-br/bioma-cerrado/informacoes, (accessed: September **2020**).
[29] Ministério do Meio Ambiente, Mata Atlântica, https://www.mma.gov.br/component/k2/item/273-mata-atl%C3%A2ntica.html, (accessed: September **2020**).
[30] G. Fusco, A. Minelli, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2010**, *365*, 547.
[31] 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, *Food Anal. Methods* **2020**, *13*, 1691.
[32] T. G. Dietterich, *Multiple Classifier Systems*, Springer, Berlin, Heidelberg **2000**, p. 1.
[33] C. Casanova, F. E. Colla-Jacques, J. G. C. Hamilton, R. P. Brazil, J. J. Shaw, *PLoS Negl. Trop. Dis.* **2015**, *9*, e0003620.
[34] O. Mangabeira, *Rev. Bras. Malariol. Doencas Trop.* **1969**, *21*, 3.
[35] A. V. Martins, A. L. Falcão, J. E. da Silva, E. S. Dias, A. V. Martins, A. L. Falcão, J. E. da Silva, E. S. Dias, *Mem. Inst. Oswaldo Cruz* **1984**, *79*, 439.
[36] R. P. Brazil, J. G. C. Hamilton, *Mem. Inst. Oswaldo Cruz* **2002**, *97*, 435.
[37] F. M. Vigoder, A. S. Araki, L. G. S. R. Bauzer, N. A. Souza, R. P. Brazil, A. A. Peixoto, *Infect. Genet. Evol.* **2010**, *10*, 734.

[38] P. E. M. Ribolla, L. T. Gushi, M. S. do Pires e Cruz, C. H. N. Costa, D. L. Costa, M. S. C. da Lima Júnior, M. E. M. C. Dorval, A. de Gutierrez Oliveira, M. F. da Cunha Santos, V. L. Fonseca Camargo-Neves, C. M. C. B. Fortaleza D. P. Alonso, *Leishmania infantum* genetic diversity and *Lutzomyia longipalpis* mitochondrial haplotypes in Brazil, https://www.hindawi.com/journals/bmri/2016/9249217/, (accessed: July **2020**).
[39] F.-J. Richard, J. H. Hunt, *Insect. Soc.* **2013**, *60*, 275.
[40] P. H. F. Shimabukuro, J. E. Tolezano, E. A. B. Galati, *Papéis Avulsos de Zoologia* **2011**, *51*, 399.
[41] J. I. Rodríguez-Fernández, C. J. B. De Carvalho, C. Pasquini, K. M. G. De Lima, M. O. Moura, G. G. C. Arízaga, *Zootaxa* **2011**, *2933*, 46.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Casaril AE, Santos CG, Marangoni BS, et al. Intraspecific differentiation of sandflies specimens by optical spectroscopy and multivariate analysis. *J. Biophotonics*. 2021;14: e202000412. https://doi.org/10.1002/jbio.202000412