Received: 17 February 2020 | Revised: 16 October 2020 | Accepted: 8 November 2020

DOI: 10.1111/jzs.12442

ORIGINAL ARTICLE

JOURNAL OF ZOOLOGICAL SYSTEMATICS AND EVOLUTIONARY RESEARCH

WILEY

Effectiveness of Fourier transform near-infrared spectroscopy spectra for species identification of anurans fixed in formaldehyde and conserved in alcohol: A new tool for integrative taxonomy

Kelly Torralvo[1] | William E. Magnusson[2] | Flávia Durgante[3,4]

[1]Programa de Pós Graduação em Ecologia, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brasil

[2]Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil

[3]Programa de Pós Graduação em Botânica, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brasil

[4]Institute of Geography and Geoecology, Department of Wetland Ecology, Karlsruhe Institute for Technology (KIT), Rastatt, Germany

Correspondence
Kelly Torralvo, Programa de Pós Graduação em Ecologia, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brasil.
Email: kellytorralvo1@gmail.com

Funding information
Fundação de Amparo à Pesquisa do Estado do Amazonas; ATTO project (German Federal Ministry of Education and Research, BMBF funds 01LB1001A); Projeto CENBAM INCT Centro de Biodiversidade Amazônica - Convênio FAPEAM/FDB/INPA nº 003/2012

Abstract

Integrative taxonomy takes into account multiple perspectives to delimit the units of diversity, which is fundamental for communication in ecological studies and conservation actions. Near-infrared spectroscopy is efficient for identification of species in many taxa and may be a viable alternative to the descriptions used in classical taxonomy as it is highly cost-effective, rapid, and can be non-destructive. Animal specimens are often conserved in alcohol and formaldehyde, which hinders some techniques. In this study, we tested the effectiveness of near-infrared spectroscopy in recognizing species of fixed specimens of closely related anuran species when the raw data were used and after preprocessing to reduce noise due to variation in surface structure. We used individuals of adult frogs belonging to different species, sizes, and time conserved in alcohol. The tests were divided into spectral readings of the belly and back of each sample in wet and predrying manipulations. The results indicated that the presence of superficial alcohol in individuals does not impede the use of the technique to recognize anuran species. We recognized anuran species using spectra collected on the back or belly with up to 100% correct identification in raw data or preprocessed data. These results open up numerous possibilities for taxonomic investigations using individuals of zoological collections worldwide, but this requires the creation of open-access data banks and availability of spectra of all species.

KEYWORDS

Amazonian frogs, near-infrared spectroscopy, species identification, taxonomic methods, zoological collections

1 | INTRODUCTION

Currently, the number of frog species is estimated at 7,248 (Frost, 2020) with 329 for the Amazon (Hoogmoed, 2019). However, it is predicted that around 40% of Neotropical species are yet to be described (Giam et al., 2012). This is corroborated by frequent descriptions of new species from the Amazon (e.g., Ferrão et al., 2017, Ferrão et al., 2018; Kaefer et al., 2019). "Species-level" identification is critical for communication in many ecological studies and conservation actions (Gotelli, 2004), but using fine taxonomic levels to

Contributing authors: William E. Magnusson (wemagnusson@gmail.com), Flávia Durgante (flaviamdflorestal@gmail.com)

measure and map biodiversity is complex and generally dependent on experts (Bortolus, 2008; Margules & Pressey, 2000).

Integrative taxonomy (Dayrat, 2005; Padial et al., 2010) takes into account multiple perspectives (e.g., phylogeography, morphology, genetics, and ecology) as complementary to delimit the units of diversity, considering conceptual and technological advances, and has been used frequently in recent studies (e.g., Funk et al., 2012; Moraes et al., 2016; Ferrão et al., 2016; Prata et al., 2018; Kaefer et al., 2019). This approach considers that taxonomic characters can be equivalent and combinable units, discarding the idea of a restrictive taxonomy for biodiversity studies (Padial et al., 2010). Mistakes in species identification can affect planning and result in inefficient management and decision-making, in some cases even directly linked to human health (Bortolus, 2008). Therefore, there is a need for simple and reliable tools suitable for different users, such as pharmacologists, physiologists, biologists, ecologists, and biodiversity managers (Dayrat, 2005).

Near-infrared spectroscopy (NIR) captures the molecular vibrations of chemical compounds in the structure of a sample (Pasquini, 2003; Pavia et al., 2010) that may be a whole organism (Vance et al., 2016) or a part of an organism (Rigby et al., 2014). From the NIR spectra, it is possible to obtain a molecular signature (Jouquet et al., 2014) attributed to refined sample recognition that functions as a physicochemical fingerprint. Fourier transform NIR (FT-NIR) is used to recover the intensities of individual wavelengths and ensure better accuracy (Pasquini, 2003).

Recent studies have shown the effectiveness of FT-NIR spectroscopy in botanical species identifications (Durgante et al., 2013; Fan et al., 2010; Hadlich et al., 2018; Lang et al., 2017; Lang et al., 2015; Prata et al., 2018). The method has also been used in vertebrate and invertebrate studies (Almeida de Azevedo et al., 2019; Rigby et al., 2014; Rodriguez-Fernandez et al., 2011; Vance et al., 2014, 2016) and may be a viable alternative to characteristics used in classical taxonomy as it is highly cost-effective, rapid, and can be non-destructive (Rodriguez-Fernandez et al., 2011; Vance et al., 2016).

The data available in collections inform or contribute to the knowledge of species biology, ecology, and biogeography (Margules & Pressey, 2000; Meineke et al., 2019). However, many species identification errors or the lack of curation of deposited samples lead to problems in data use. Readings of the spectra in dry plants of herbarium collections brought an alternative to investigate identification problems using specimens already available without the need for new collections (Durgante et al., 2013; Fan et al., 2010; Lang et al., 2017; Lang et al., 2015; Prata et al., 2018). The same limitations found in botanical collections apply to zoological collections (Meineke et al., 2019). Unlike plants, animal specimens are often conserved in alcohol and formaldehyde, which hinders some techniques that use DNA extraction (Friedman & DeSalle, 2008; Jaksch et al., 2016; Schander & Halanych, 2005) and may also be limiting for using NIR techniques (Almeida de Azevedo et al., 2019; Pasquini, 2003).

Fourier transform-NIR is affected both by the chemical composition of the sample and by the physical structure of the surface (Pasquini, 2003, 2018). Therefore, readings may differ between different body regions on the same individual because of differences in rugosity. To reduce the effect of physical differences in samples, FT-NIR spectra can be preprocessed, which sometimes allows better detection of chemical differences (Pasquini, 2018; Rinnan et al., 2009). Therefore, we analyzed the data before and after preprocessing using the Savitzky-Golay method (Barak, 1995; Rinnan et al., 2009).

The efficient species identification of anuran species in biological collections by FT-NIR spectra is a promising new taxonomic method that may allow the construction of a reference database for species recognition. However, first it is necessary to determine whether the method can distinguish closely related species based on alcohol- and formaldehyde-fixed specimens.

Considering the potential of FT-NIR spectroscopy and the need to develop tools for integrative anuran taxonomy, our principal objective was to test the effectiveness of the technique in recognizing anuran species with specimens that had been fixed with formaldehyde and stored in alcohol. We also considered two methodological questions: (a) is it necessary to dry the alcohol on the surface of samples and (b) what position on the body of individuals (back or belly) gives a more efficient spectral reading to distinguish anuran species?

2 | MATERIAL AND METHODS

2.1 | Sample collection and FT-NIR spectroscopy readings

All anurans used were in the collection of amphibians and reptiles of the National Institute of Amazonian Research—INPA, and had been fixed in formaldehyde and maintained in 70% alcohol for periods ranging from 3 to 28 years (Appendix 1). Pairs (and one trio) of closely related and sympatric species from four families were used. We used sympatric individuals because practical questions are often about species assemblages and to reduce the probability of including cryptic species. Many species of Amazonian frogs that previously were thought to have wide distributions have recently been separated into geographically separated species (e.g., Peloso et al., 2018; and Carvalho et al., 2020).

Two sets of readings were used in this study, one for each series of tests undertaken. Some species differed between the two series because some specimens were not available for loan on both occasions, but 53 individuals were used in both series. The first series of tests evaluated the influence of alcohol drying on species identification through the FT-NIR spectrum. For this series, a total of 89 anuran individuals belonging to eight species were used. Species of different body sizes (average 1.4–10.3 cm long and 0.5–3.2 cm wide) and storage time in alcohol (3–28 years) were selected (Table 1) to maximize variation and test the tool on realitistic samples from collections. Measurements in this dataset were divided into spectral readings of the belly and back of each sample (Figure 1) in two manipulations:

TABLE 1 Characteristics of samples used to test the pre-reading drying effect on species identification

Species	No.	Body length[a]	Body width[a]	Age[b]	Drying minutes
Adenomera andreae (Müller, 1923)	13	1.93	0.68	6 and 3	30–40
Adenomera hylaedactyla (Cope, 1868)	11	2.30	0.86	27 and 12	20–30
Leptodactylus fuscus (Schneider, 1799)	17	4.40	2.0	26	25–35
Leptodactylus longirostris Boulenger, 1882	10	3.6	1.4	25 and 13	20–30
Callimedusa tomopterna (Cope, 1868)	10	4.20	0.9	23 and 7	20–30
Phyllomedusa bicolor (Boddaert, 1772)	9	10.3	3.2	28 and 24	20–30
Amazophrynella minuta (Melin, 1941)	9	1.40	0.5	15	20–30
Amazophrynella manaos Rojas-Zamora, Carvalho, Ávila, Farias, and Hrbek, 2014	10	2.20	0.7	Unknown	20–30

[a]Mean snout-urostyle length of individuals in cm.

[b]Storage time in alcohol in years.

Wet—excess alcohol dried with an absorbent cloth (reusable cloth—Scott Duramax®) and spectral reading;

Dry—The same animals were used as in the wet manipulation, but a second reading was taken after the specimen had been drying at 24°C for 20 to 40 min (Table 1).

The drying manipulation did not eliminate the presence of formaldehyde or alcohol in the samples and was used to evaluate the need for removal of liquid preservative from the surfaces of specimens to use the technique.

For the second data series, used to identify the position on the body of the frogs with the most effective spectrum for recognizing species, we used a set of 100 samples of adult frogs belonging to nine species with average lengths ranging from 1.4 to 10.3 cm and storage time in alcohol from 8 to 28 years (Table 2), with no predrying. Spectral readings were taken at different sites on the belly and back of each specimen (Figure 1).

The spectral readings were taken with an Antaris II FT-NIR Analyzer provided by the botany spectroscopy laboratory of the National Institute of Amazonian Research. The equipment was programmed to near-infrared spectral readings of 16, with 8 cm^{-1} resolution and wavenumber in the range of 10,000 to 4,000 cm^{-1}, which corresponds to the electromagnetic spectrum in wavelengths from 1,000 to 2,500 nm. Each spectrum consisted of 1,557 individual absorbance values. A black rubbery plate (E.V.A.—Ethylene Vinyl Acetate) was placed on the smaller individuals to prevent light scattering since these samples did not cover the entire reading area. The background was calibrated automatically by the equipment every 4 hr. The spectra of each individual were checked in order to detect anomalies that would be excluded in the reading procedure before analysis (see Figure S1).

2.2 | Data analysis

We used the first data series in different combinations between reading position (back and belly—Figure 1a) and alcohol drying time. In the second series, we used the mean of the spectral values for dorsal and ventral position and original spectra values at randomly selected sites on the back and belly (Figure 1b). Different combinations of data were used to test different reading models (Table 3).

For a preview of the spectral behavior, the raw and preprocessed data were subjected to a principal component analysis—PCA. To test the potential of the technique to recognize anuran species in different data series, the models were subjected to discriminant analysis (LDA) and evaluated through two forms of cross-validation: holdout method (70–30) with 100 randomizations that uses 70% of the data to calibrate the model and 30% (independent group) to test the model; and the leave-one-out (LOO) method, which tests sample by sample by removing one of the n specimens at a time (independent group) and testing its identification with the model generated with n-1 specimens.

The statistical analyses were done in the R statistical program (R Core Team, 2013) with the support of the MASS package (Venables & Ripley, 2002) and the klaR package (Weihs et al., 2005), with adaptations of command scripts already used in data processing and spectral analysis of botanical samples (available in: http://www.botanicaamazonica.wiki.br/labotam/doku.php?id=analises:nir:inicio).

We used the Savitzky-Golay filter to pre-process the spectra through the *savitzkyGolay* function available in the PROSPECTR package (Stevens et al., 2020). This function requires the provision of three arguments: polynomial order (p), differentiation order (m), and window size (w). We created combinations of these values (p from 2 to 10; m from 0 to 2; w from 3 to 101, in regular intervals of 2 units, and with the value of w always greater than the order of the polynomial within the combination) and executed for each combination of filtering of the raw spectra followed by an LDA holdout, in which we reserved 70% of the data for training the model and 30% for testing the model. The data were scaled and centralized before executing each LDA. At the end of the execution, we calculated the performance of each filtered set by combining the values of p, m, and w with the *MulticlassSummary* function of the CARET package (Kuhn, 2012). We selected the best Savitzky-Golay filter for the four models of the first series of tests and for the three models of spectral-mean of the second series of tests, considering the sensitivity values returned by the function and we analyzed preprocessed data using holdout method (70–30) with 100 randomizations and the

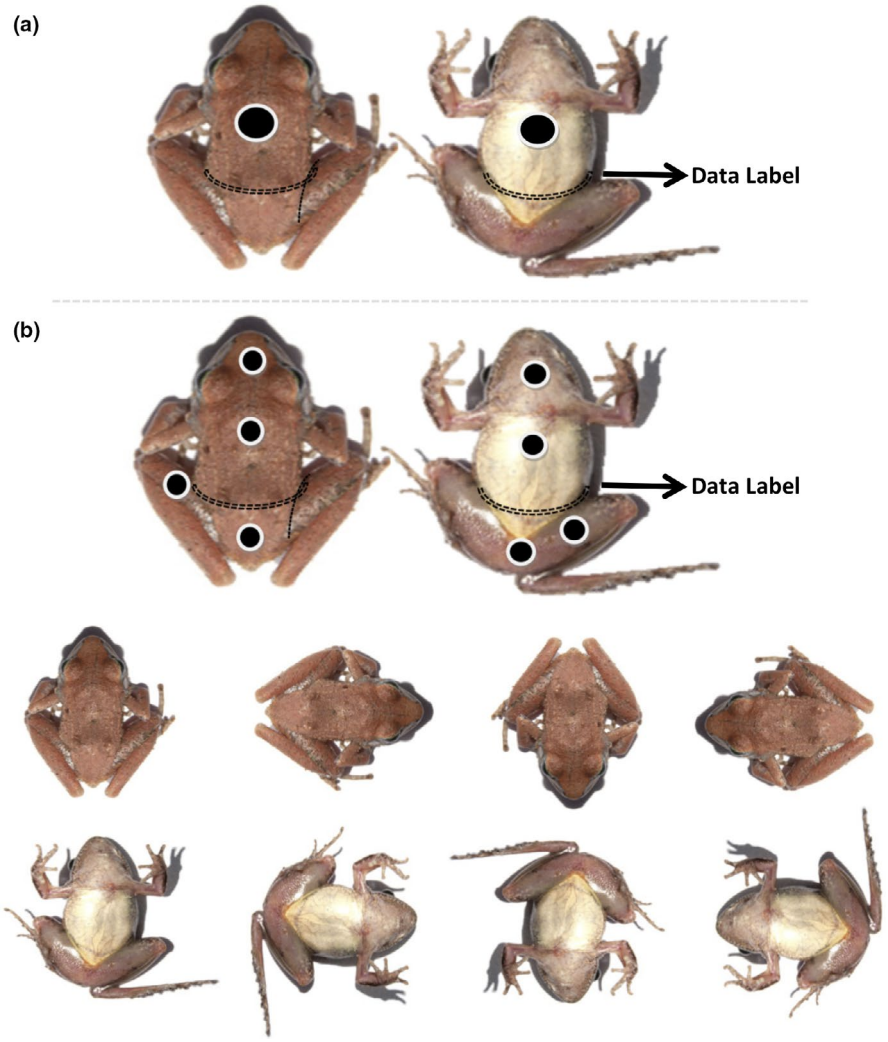


FIGURE 1 Fourier transform near-infrared spectroscopy reading sites on the body of anurans. (a) Readings at central points on the back and belly, assigned to all samples for series one; (b) Readings at eight different sites on the back and belly, depending on the sizes of frogs for series two. For larger-sized frogs: on the back—1. front, close to the snout and between the eyes; 2. central, between the arms; 3. posterior, between the legs; 4. outer thigh; on the belly—1. front, on the throat; 2. central, between the arms; 3. posterior, between the legs; 4. inner thigh. For smaller individuals, eight different readings were taken from center of the body, with 90° turns of the sample on the reader in back and belly positions

TABLE 2 Characteristics of the samples used to test the best spectral-reading site for species identification

Species	No.	Body length[a]	Body width[a]	Age[b]
Amazophrynella manaos Rojas-Zamora, Carvalho, Ávila, Farias, and Hrbek, 2014	12	2.2	0.7	Unknown
Amazophrynella bilinguis Kaefer, Rojas-Zamora, Ferrão, Farias, and Lima, 2019	12	1.9	0.5	8
Leptodactylus fuscus (Schneider, 1799)	12	4.7	2.1	26
Leptodactylus longirostris Boulenger, 1882	10	3.6	1.4	25 and 13
Callimedusa tomopterna (Cope, 1868)	10	4.2	0.9	23 and 7
Phyllomedusa bicolor (Boddaert, 1772)	9	10.3	3.2	28 and 24
Allobates caeruleodactylus (Lima and Caldwell, 2001)	9	1.6	0.4	22
Allobates magnussoni Lima, Simões, and Kaefer, 2014	13	1.8	0.5	12
Allobates tapajos Lima, Simões, and Kaefer, 2015	13	1.6	0.5	12

[a]Mean snout-urostyle length of individuals in cm.

[b]Storage time in alcohol in years.

TABLE 3 Description of spectral models tested for anuran species identification

Combinations	Descriptions
MODELS FOR SERIES 1	
MODEL 1.1: Back – wet	Spectra collected on the back without alcohol drying of samples
MODEL 2.1: Back – dry	Spectra collected on the back with alcohol drying of samples
MODEL 3.1: Belly – wet	Spectra collected on the belly without alcohol drying of samples
MODEL 4.1: Belly – dry	Spectra collected on the belly with alcohol drying of samples
MODELS FOR SERIES 2 (all wet)	
MODEL 1.2: Mean Back	Mean spectra collected on the back of individuals
MODEL 2.2: Mean Belly	Mean spectra collected on the belly of individuals
MODEL 3.2: Mean Back and Belly	Mean spectra collected on the back and belly of individuals
MODEL 4.2: Back – random all spots	All spectra collected at randomly selected points on the back of individuals
MODEL 5.2: Belly – random all spots	All spectra collected on the at randomly selected points on the belly of individuals
MODEL 6.2: Random all data	All spectra collected at randomly selected points on both the back and belly of individuals

leave-one-out (LOO) method to calculate the species prediction-accuracy rates (%).

For the raw and preprocessed spectral-mean data of the back and belly, used in the second series, we also tested the species prediction-accuracy rates with a reduction in the number of spectral absorbance values (stepwise—most informative regions of the spectrum), through LDA holdout. The data were scaled and centralized before executing each LDA.

We also analyzed the influence of the size and age of the samples on species identification errors of the raw and preprocessed data for all condensed models, using a generalized linear model (GLM), with binomial family.

3 | RESULTS

3.1 | Influence of alcohol drying on species identification

Using the preprocessed spectra, the maximum sensitivity value for the first series of tests was achieved with the Savitzky-Golay filter in the first derivative, with orders of polynomials 2 and 3 and windows between 25 and 97 (Table 4).

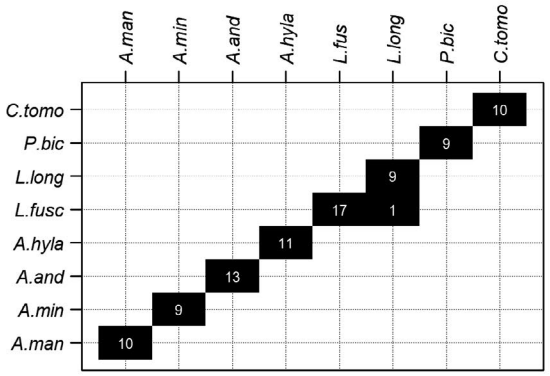
Principal components analysis (PCA) was used for an initial visual inspection of the data. The first axes of the PCAs on preprocessed data explained between 37.4% and 42.8%, and the second axes explained 24.6%–30.1% of the total variability. Similar results were obtained for the raw data, 57% and 89% in first axes, and the second axes explained 8% to 41% of the total variability. The ordinations grouped measures taken on the same part of the body (back or belly), but individuals of both manipulations (dry and wet) were dispersed throughout the ordination space (Figures S2 and S3).

To estimate the error rate in classification, we used LDA. Species were more than 92% correctly identified in all datasets testing the influence of excess alcohol. The belly-wet model achieved better recognition (LOO – 100%) with both raw data and preprocessed data. Preprocessing resulted in few changes, sometimes increasing and sometimes decreasing the discrimination power of the model (Table 4), but overall species discrimination was similar for raw and preprocessed data. The LOO validation results for raw data in the back-wet model readings misidentified an individual of *Leptodactylus*

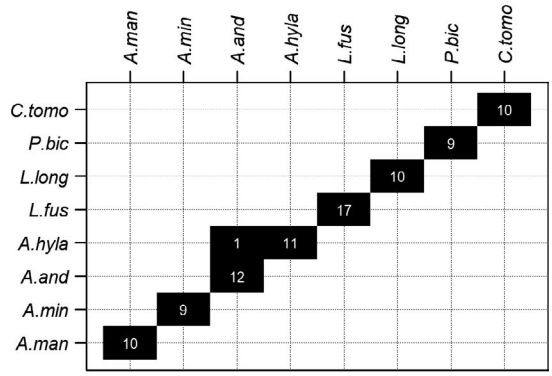
TABLE 4 Species prediction-accuracy rates measured in discriminant analysis with holdout and leave-one-out (LOO) validations, in different combinations of position and drying for raw and preprocessed data in (test series 1)

	Raw data					Preprocessed data							
	Holdout (%)				LOO	Savitzky-Golay Filter			Holdout				LOO
Combinations	Mean	Min	Max	CI (99%)	(%)	P	M	W	Mean	Min	Max	CI (99%)	(%)
MODEL 1.1: Back – wet	93.9	83.3	100	92.63–95.28	98.9	3	1	33	92.1	75	100	90.6–93.0	98.9
MODEL 2.1: Back – dry	95.5	87.5	100	94.50–96.49	96.6	2	1	97	98.3	87.5	100	97.6–99.1	100
MODEL 3.1: Belly – wet	98.9	87.5	100	98.32–99.51	100	2	1	33	98.7	87.5	100	98.0–99.3	100
MODEL 4.1: Belly – dry	98.0	87.5	100	97.39–98.76	98.9	2	1	25	96.1	87.5	100	95.2–96.9	96.6

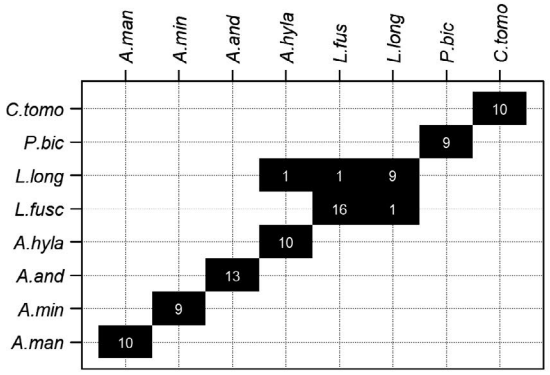
Model 1.1 - Back wet- raw data



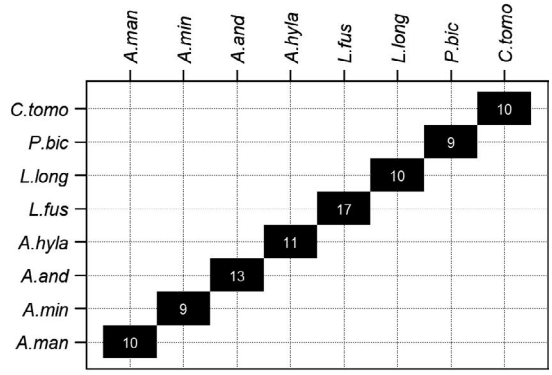
Model 1.1 - Back wet- preprocessing



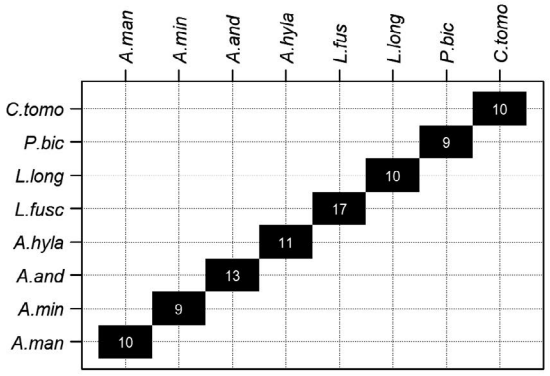
Model 2.1 - Back dry- raw data



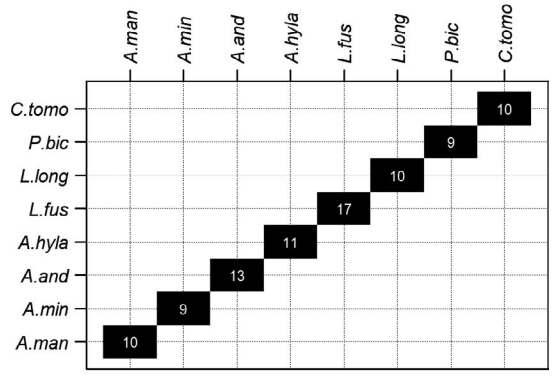
Model 2.1 - Back dry- preprocessing



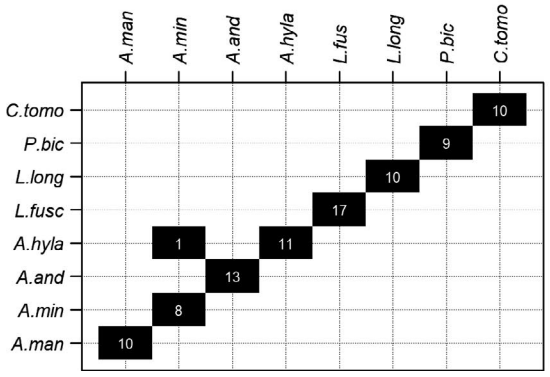
Model 3.1 - Belly wet- raw data



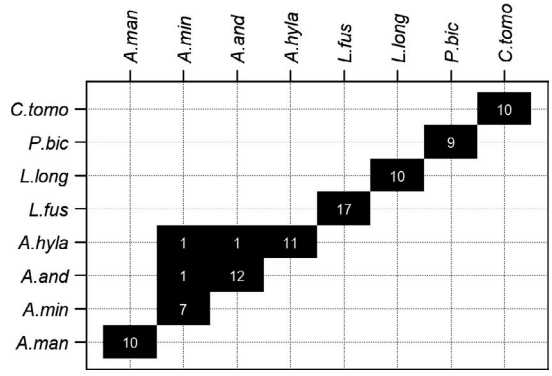
Model 3.1 - Belly wet- preprocessing



Model 4.1 - Belly dry- raw data



Model 4.1 - Belly dry- preprocessing



longirostris as *Leptodactylus fuscus* and for preprocessed data, a *Adenomera hylaedactyla* was identified as a *A. andreae*. The same species were confused with each other in the back-dry model, but with different individuals and no errors for preprocessed data. In the belly-dry model for raw and preprocessed data, there were errors in the identification for the species *A. andrea*e, *A. hylaedactyla*, and *Amazophrynella minuta* (Figure 2).

3.2 | Effect of body site on species identification

The preprocessed spectra of the three mean models in the second test series also returned the maximum sensitivity value through the Savitzky-Golay filter in the first derivative, with orders of polynomials 2 and windows between 9 and 39 (Table 5).

The first axes of the PCAs based on preprocessed data explained 36.4%–41.7% and the second axes explaining 26.1%–34.3% of the total variance. Similar results were obtained with the raw data with 66%–95% of explained variance in the first axes, and the second axes explaining 5%–32% of the total variability. Most individuals of each species were grouped for the readings with mean spectra of positions (back and belly) tested (Figures S4 and S5).

Results from the second series based on raw data also had high hit rates for the three spectral-mean models tested (mean rates between 94.4% and 98.8% with holdout, and 97% to 100% with LOO—Table 5). As with data from the first series, the model constructed with the means of the belly spectral readings resulted in 100% correct prediction of the species. Preprocessing did not increase the ability of the LDA to distinguish species in models of the second data series; the rates remained high and similar (Table 5).

The LDA-LOO result matrices with back spectral means resulted in an error for two individuals of *L. fuscus* identified as *L. longirostris* for raw and preprocessed data, and errors in the identification of *Amazophrynella bilinguis* and *Allobates tapajos*. There were no errors when using the mean-belly model for raw data, but for preprocessed data one individual of *A. bilinguis* was identified as *Allobates caeruleodactylus*. With mean spectra of back and belly of raw data, an *L. fuscus* individual was identified as *L. longirostris* and an *Al. tapajos* individual was identified as *A. bilinguis*. For the preprocessed data, two individuals of *L. fuscus* was identified as *L. longirostris* (Figure 3).

Low prediction success occurred in the models with a randomly chosen single raw spectrum from the back (81.4%) and belly (88.5%) analyzed separately that we tested to understand the performance with different body sites. For the model tested with the single raw spectrum randomly selected from all readings (back and belly together) at different body sites (eight sites- Figure 1), the mean correct-prediction rates were even lower. With these data, preprocessing resulted in an increase of about 10% in the correct-prediction rates (raw data 69.2% in holdout and 73% in LOO/preprocessed data 78.4% in holdout and 83% in LOO). For this analysis, the Savitzky-Golay filter in the second derivative, with polynomial order 4 and window 91 was selected as the best combination (Table 5).

Variable-selection (stepwise) analysis in unpreprocessed mean back and belly models (models 1.2 and 2.2) resulted in 33 spectral regions that best distinguished species. By reducing the number of variables from 1,557 to 33, the predictive power decreased, but was still high. The holdout validation results using the selected variables from the models with spectral means were 88.3% (CI, 86.8–89.8) of hits for the back model and 90.1% (CI, 88.6–91.6) for the belly model. The most informative spectral regions to recognize the species were distinct between the back and the belly and were mainly distributed in the spectral range of 4,000 to 7,100 cm^{-1} for both back and belly and additionally in regions 9,000 to 10,000 cm^{-1} for the belly (Figure 4).

For the preprocessed mean back and belly models (models 1.2 and 2.2), stepwise selection of variables resulted in 33 more informative areas of the spectrum for both models. The most informative regions were in the spectral range between 4,072 and 8,554 cm^{-1} with several points between 4,072 and 6,444 cm^{-1} and in the mean-belly model a distribution between 4,018 and 8,543 cm^{-1} with several points between 4,018 and 4,674 cm^{-1} (Figure 5). With reduction in the number of variables, the predictive power was lower for both models, but results were similar for raw and preprocessed data. The results of the holdout validation using the variables selected from the models with preprocessed mean spectra were 94.2% (CI, 92.9–95.5) of prediction accuracy in the dorsal model and 93% (CI, 91.7–94.3) in the belly model. The complete list of selected variables for each model for raw and preprocessed data is given in the Tables S1 and S2.

The size and age of the samples had no significant influence on species identification errors with the preprocessed (Body length – $p = .11$; Body width – $p = .11$; Age – $p = .58$) or raw data (Body length – $p = .11$; Body width – $p = .15$; Age – $p = .08$).

4 | DISCUSSION

The different models tested indicate that it is possible to generate anuran spectral models to recognize species even in alcohol-conserved samples of zoological collections. The presence of residual alcohol on individuals does not impede the efficient use of the technique to recognize anuran species, as predicted by other studies (Almeida de Azevedo et al., 2019; Pasquini, 2018). All individuals used in this

FIGURE 2 Linear discriminant analysis-leave-one-out results matrix (test series 1) for raw and preprocessed data. The species names in the calibration are given in rows, while predicted names are given in columns. Diagonal values are correct predictions, and off-diagonal values are incorrect predictions. The number within the squares refers to the number of samples used in the models for each species. *A.and = Adenomera andreae*, *A.hyla = Adenomera hylaedactyla*, *L.fusc = Leptodactylus fuscus*, *L.long = Leptodactylus longirostris*, *C.tomo = Callimedusa tomopterna*, *P.bic = Phyllomedusa bicolor*, *A.min = Amazophrynella minuta*, *A.man = Amazophrynella manaos*

TABLE 5 Species prediction-accuracy rates measured in discriminant analysis with holdout and leave-one-out (LOO) validations, in different combinations of position for raw and preprocessed data (test series 2)

	Raw data					Preprocessed data							
	Holdout (%)				LOO	Savitzky-Golay Filter			Holdout				LOO
Combinations	Mean	Min	Max	CI (99%)	(%)	P	M	W	Mean	Min	Max	CI (99%)	(%)
MODEL 1.2: Mean Back	95.6	84	100	94.5–96.7	97	2	1	39	96.0	84	100	95.0–96.9	97
MODEL 2.2: Mean Belly	98.8	96	100	98.3–99.2	100	2	1	11	96.6	84	100	95.7–97.5	99
MODEL 3.2: Mean Back and Belly	94.4	84	100	93.2–95.7	99	2	1	9	97.2	88	100	96.5–97.9	98
MODEL 4.2: Back – random all spots	81.4	60	96	79.6–83.2	85	—	—	—	—	—	—	—	—
MODEL 5.2: Belly – random all spots	88.5	76	100	87.2–89.8	93	—	—	—	—	—	—	—	—
MODEL 6.2: Random all data	69.3	48	96	67.1–71.5	73	4	2	91	78.4	60	92	76.5–80.3	83

study contained traces of the alcohol in which they are preserved. All or most of the samples coming from collections will contain traces of formaldehyde and alcohol because specimens are almost always conserved in alcohol after being fixed in formaldehyde before they are deposited. Therefore, we believe that the presence of alcohol and formaldehyde did not affect discrimination because they were common to all individuals in the collection (Simmons, 2002). In addition, most anurans included in the tests were correctly identified, and errors were independent of fixation time, which varied from 3 to 28 years, and independent of individual size, which average varied from 1.4 to 10.3 cm long and 0.5 to 3.2 cm wide.

Both series of tests we carried out with spectra that were not preprocessed indicated higher effectiveness when the spectral reading was done on the belly. Models with the mean-belly spectra always had higher predictive power for raw data. Belly readings in live anurans have already been used to identify species and sex (Vance et al., 2014), but no test of the reading position on the anuran's body was presented. We believe that the greater efficiency in species identification with the belly spectrum is because it is a more uniform location and the contact with the equipment reading area was more regular than for the back spectrum (see Figure S1). Another explanation may be related to the minor variation in color, roughness, and environmental factors (e.g., fungi) that may be present on the back of individuals. Preprocessing data is suggested to smooth the sample physical variation on the spectra and was expected to improve LDA model performance. However, in this study, preprocessing smoothed the differences among readings from different parts of the body (Data S1), but it did not improve the performance of most models and it was not necessary to obtain high hit rates in species identification.

We recommend the identification of the position on the anuran body where the NIR spectral reading was made to ensure a standard method for rapid and practical species identification. However, tests with the back spectrum also showed relatively high rates of species identification accuracy, indicating that it is possible to capture important information also at this body position and these could be used if the specimens had damaged bellies. As studies with FT-NIR may include individuals with only some body parts (e.g., Rigby et al., 2014), we suggest that the spectral readings be collected at different back and belly sites to construct a more complete spectral-reference bank.

In this study, preprocessing increased correct-prediction rates over raw spectra when samples were from random body sites. This is probably related to the efficiency of smoothing physical differences from the readings that were made in very different sites (e.g., on the throat and inner thigh—Figure 1) used in the same model. Therefore, we recommend preprocessing in similar cases, to obtain maximum accuracy when it is not possible to standardize body sites.

The high level of accuracy maintained even with the reduction in the number of preprocessed spectral variables from 1,557 to 33 in the tests indicates the robustness of the spectral data in recognizing anuran species with few spectral bands. This selection in addition to reducing the number of variables to less than 1/3 the number of individuals, as recommended by Williams and Titus (1988), decreases a possible collinearity interference of spectral data in discriminant analysis. The selection of variables also helps to recognize the regions of the electromagnetic spectrum that are important in the identification of anuran species and this may be important when data are analyzed with spectrophotometers with different spectral ranges.

For raw data, spectral regions selected as the most informative to distinguish species across the back spectrum occurred from 3,999 to 7,169 cm^{-1} with several points ranging from 4,246 to 4,786 cm^{-1}. In this region, there are bands that express signals of functional groups, such as methylenes (4,261 cm^{-1}), aromatics (4,246 cm^{-1}), proteins (4,265 cm^{-1}), and polyamides (4,701 cm^{-1}) (Workman & Weyer, 2008). On the belly, the most informative spectral regions occurred in a wider region of the electromagnetic spectrum, occurring from 4,011 to 9,993 cm^{-1}. Nitrogen functional

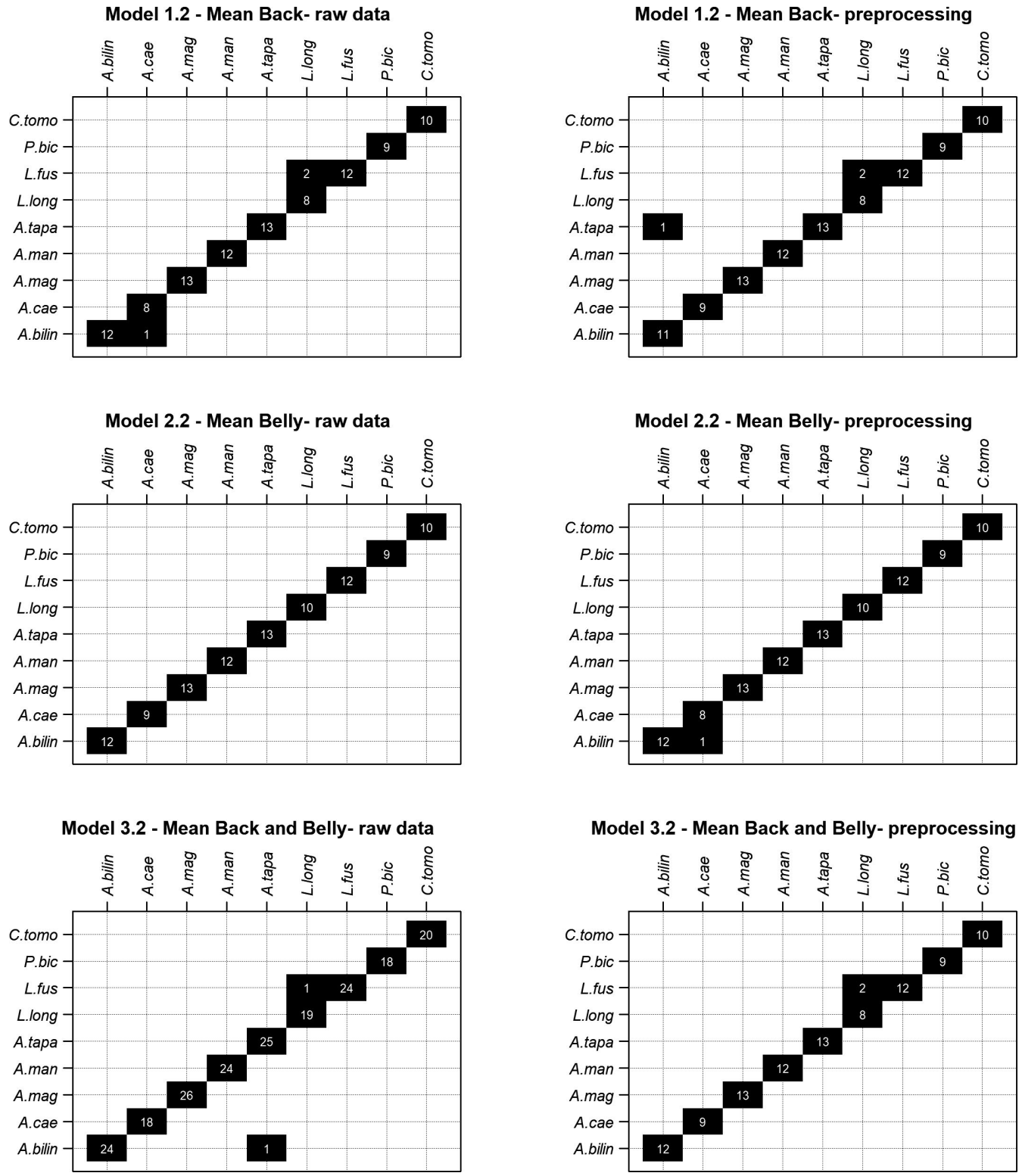


FIGURE 3 Linear discriminant analysis-leave-one-out results matrix for the tests with spectral-mean models (test series 2) for raw and preprocessed data. The species names in the calibration are given in rows, while predicted names are given in columns. Diagonal values are correct predictions, and off-diagonal values are incorrect predictions. The number within the squares refers to the number of samples used in the models for each species. *A.bilin* = *Amazophrynella bilinguis*, *A.cae* = *Allobates caeruleodactylus*, *A.mag* = *Allobates magnussoni*, *A.man* = *Amazophrynella manaos*, *A.tapa* = *Allobates tapajos*, *L.long* = *Leptodactylus longirostris*, *L.fus* = *Leptodactylus fuscus*, *P.bic* = *Phyllomedusa bicolor*, *C.tomo* = *Callimedusa tomopterna*

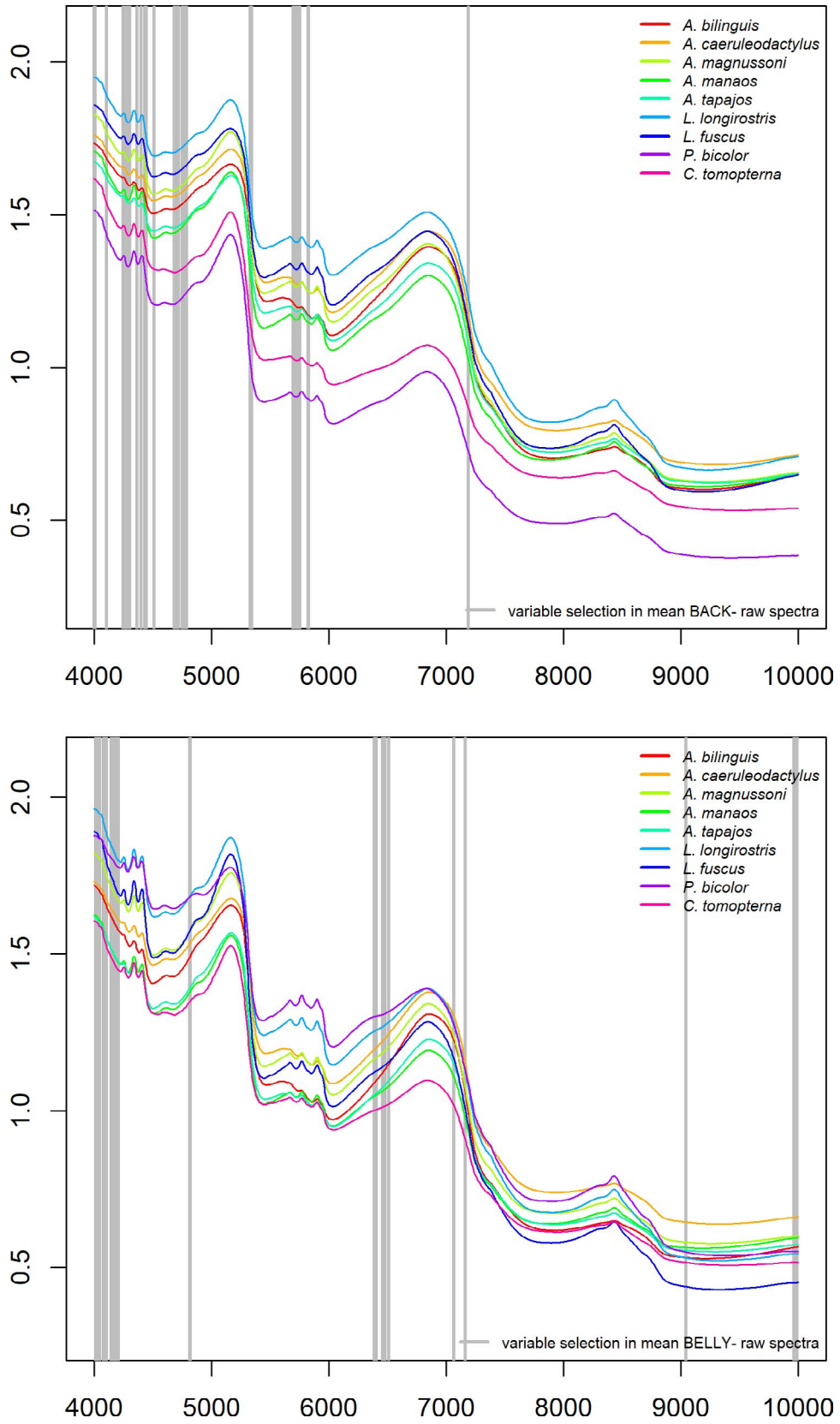


FIGURE 4 Spectra with the most informative regions selected by the stepwise method for the back and belly spectral-mean models with raw data—test series 2

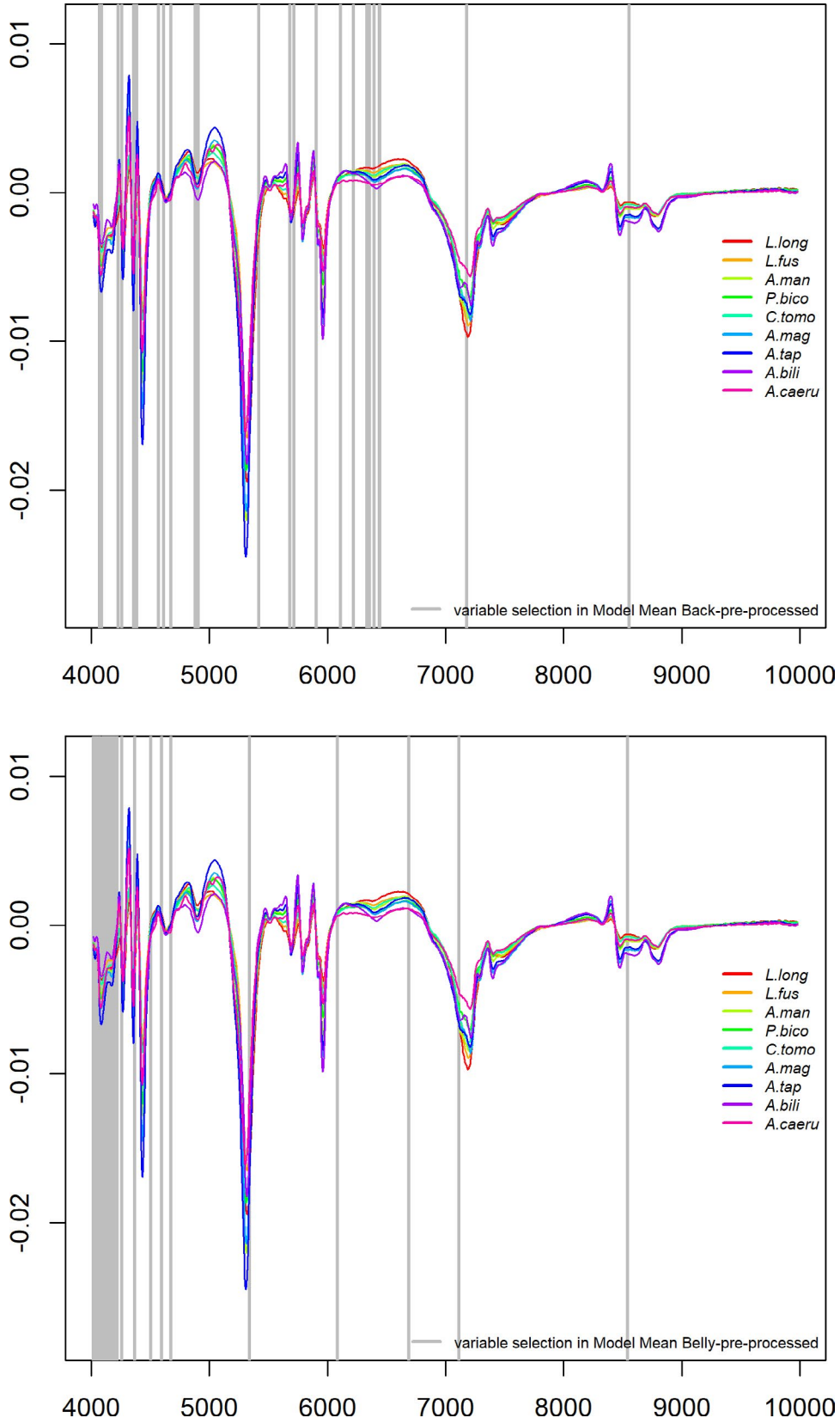


FIGURE 5 Spectra with the most informative regions selected by the stepwise method for the back and belly spectral-mean models with preprocessing—test series 2

groups characterized as starches, proteins, and hydrocarbons are recognized among the selected spectral bands (4,049 cm^{-1}), as well as alkanes and cycloalkanes (7,065 and 7,162 cm^{-1}) and O-H bonded chemical compounds (4,184 cm^{-1}) (Workman and Weyer, 2008). For preprocessed data, the spectral regions selected as the most informative to distinguish species in both mean-spectra models ranged between 4,018 and 8,554 cm^{-1} with several points between 4,072 and 6,444 cm^{-1} in the back mean model and 4,018 and 4,674 cm^{-1} in the belly mean model. In this region, there are bands that express signals of functional groups, such as methylenes (4,261, 5,708, 5,731, and 5,754 cm^{-1}), aromatics (4,080 and 4,246 cm^{-1}), proteins (4,265 cm^{-1}), and polyamides (4,701 cm^{-1}) (Workman & Weyer, 2008). In addition, nitrogen functional groups characterized as starches, proteins, and hydrocarbons are recognized among the selected spectral bands (4,049 cm^{-1}), as well as alkanes and cycloalkanes (7,065 and 7,162 cm^{-1}) and O–H bonded chemical compounds (4,184 cm^{-1}) (Workman & Weyer, 2008). The region of O–H bonds of alcohol with a peak at 7,090 cm^{-1} was not selected as the most informative to recognize species, corroborating the conclusion that alcohol presence in individuals does not impede the use of the technique. The regions between 4,760 to 4,445cm^{-1} contain the aldehyde groups in which formaldehyde is expressed (Workman and Weyer, 2008), and these were selected as informative in some selections in the back model raw data and in both models with preprocessing. Possibly, fixation produces useful chemicals to distinguish species, but this should not be a problem unless analyses mix fixed and fresh specimens.

The species prediction-accuracy rates obtained with the both raw and preprocessed data spectra reinforce the efficiency of the method in the identification of species, even with the presence of alcohol and formaldehyde in individuals. Preprocessing may control noise related to physical differences among samples that were captured by the spectrum and an adjustment of the baseline (Pasquini, 2018; Rinnan et al., 2009), though this had little effect on the ability to correctly identify specimens in this study. In addition, due to the high species prediction-accuracy rates in the spectral-mean models of the back and belly, we suggest that both anuran body positions can be used successfully in species identification. However, the reduced accuracy when random positions were used indicates that body position should be controlled or that body position should be included as a variable in the LDAs using the mean of all body positions.

In general, the results presented for both series of tests indicate the effectiveness of the use of FT-NIR spectra in the identification of anuran species fixed in formaldehyde and conserved in alcohol, reaching up to 100% accuracy. These results open up numerous possibilities for investigations using samples that are in zoological collections worldwide and are not restricted to species identification. The identification of caste, sex, and age has already been tested using spectral data in anurans and other animal groups (e.g., Almeida de Azevedo et al., 2019; Rigby et al., 2014; Vance et al., 2014).

We only investigated the efficiency of the FT-NIR technique for the identification of anuran species from the zoological collections into species that have already been described. However, the technique can also be used to identify groups of individuals that do not conform to an already determined group/class (e.g., soft independent modeling of class analogy—SIMCA) (Pasquini, 2018), and FT-NIR could also be used to identify the presence of undescribed species.

The number of samples/individuals of each group/species will be decisive for the effectiveness of the analysis. The FT-NIR technique returns predictions based on the reported reference group (Burns & Ciurczak, 2007; Pasquini, 2003, 2018), so the greater the number and diversity of samples, the greater the representativeness of each group (here species) and the chance of correct prediction. We have shown that the technique can distinguish closely related species in sympatry. However, within-species geographic variation results in phenotypical variation due to the interaction between individuals and the environment, and these signals may be captured by spectra. Therefore, the technique will be most effective for species identification if the discriminant functions are based on specimens from across the species range.

Considering that spectroscopy returns a unique physicochemical fingerprint for each sample, the technique represents further evidence for taxonomic classification. This integration of spectra into taxonomy has recently been used in botanical studies (Prata et al., 2018). However, for the efficient use of FT-NIR spectra in species identification, there is a need for reference spectral data that can calibrate the models that will be used to classify new samples (Pasquini, 2003). Additionally, to expand the use of the technique by different research groups, it is important to understand the spectra generated by different instruments when defining the inclusion of reference data that will be made available. This will only be possible when the spectra are linked to standardized taxonomic databases and made available with open access (Miralles et al., 2020). The spectra collected in this study are available at: https://ppbiodata.inpa.gov.br/metacatui/view/PPBioAmOc.592.

ACKNOWLEDGEMENTS

We thank Albertina Lima and Miquéias Ferrão for help in identifying the species used, Ariane Silva for providing samples, Mike Hopkins and Alberto Vicentini for the loan of the equipment, Celio Pasquini and Kelly M. Narimoto for assistance with the use of the equipment and clarifications about the technique. Ricardo Perdiz, Pedro Pequeno, and Flávio Costa assisted with the analyses, and Thiago Bicudo helped produce the images. Financial support was provided by the Fundação de Amparo a Pesquisa do Amazonas—FAPEAM, ATTO project (German Federal Ministry of Education and Research, BMBF funds 01LB1001A) and Projeto CENBAM INCT Centro de Biodiversidade Amazônica - Convênio FAPEAM/FDB/INPA nº 003/2012. Data management and scientific supported was facilitated through the Programa de Pesquisas em Biodiversidade (PPBio) and Ecologia, monitoramento e uso sustentável de áreas úmidas (MAUA) group.

ORCID

Kelly Torralvo https://orcid.org/0000-0001-8089-7652
William E. Magnusson https://orcid.org/0000-0003-1988-3950
Flávia Durgante https://orcid.org/0000-0002-5517-8821

REFERENCES

Almeida de Azevedo, R., Wellington de Morais, J., Lang, C., & de Sales, D. C. (2019). Discrimination of termite species using Near-Infrared Spectroscopy (NIRS). *European Journal of Soil Biology*, *93*, 103084.

Barak, P. (1995). Smoothing and Differentiation by an Adaptive-Degree Polynomial Filter. *Analytical Chemistry*, *67*, 2758–2762.

Bortolus, A. (2008). Error cascades in the biological sciences: The unwanted consequences of using bad taxonomy in ecology. *Ambio*, *37*, 114–118.

Burns, D. A., & Ciurczak, E. W. (Eds.). (2007). *Handbook of near-infrared analysis*. CRC press.

Carvalho, T. R., Moraes, L. J., Lima, A. P., Fouquet, A., Peloso, P. L., Pavan, D., Drummond, L. O., Rodrigues, M. T., Giaretta, A. A., Gordo, M., Neckel-Oliveira, S., & Haddad, C. F. B. (2020). Systematics and historical biogeography of Neotropical foam-nesting frogs of the *Adenomera heyeri* clade (Leptodactylidae), with the description of six new Amazonian species. *Zoological Journal of the Linnean Society*, 1–39. [Epub ahead of print].

Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, *85*, 407–415.

Durgante, F. M., Higuchi, N., Almeida, A., & Vicentini, A. (2013). Species spectral signature : discriminating closely related plant species in the amazon with near-infrared leaf-spectroscopy. *Forest Ecology and Management*, *291*, 240–248.

Fan, Q., Wang, Y., Sun, P., Liu, S., & Li, Y. (2010). Discrimination of Ephedra plants with diffuse reflectance FT-NIRS and multivariate analysis. *Talanta*, *80*, 1245–1250.

Ferrão, M., Colatreli, O., de Fraga, R., Kaefer, I. L., Moravec, J., & Lima, A. P. (2016). High species richness of scinax treefrogs (Hylidae) in a threatened Amazonian landscape revealed by an integrative approach (WJ Etges, Ed.). *PLoS One*, *11*, e0165679.

Ferrão, M., Moravec, J., de Fraga, R., de Almeida, A. P., Kaefer, I. L., & Lima, A. P. (2017). A new species of Scinax from the Purus-Madeira interfluve, Brazilian Amazonia (Anura, Hylidae). *ZooKeys*, *706*, 137–162.

Ferrão, M., Moravec, J., Kaefer, I. L., de Fraga, R., & Lima, A. P. (2018). New species of scinax (Anura: Hylidae) with red-striped eyes from Brazilian Amazonia. *Journal of Herpetology*, *52*, 472–488.

Friedman, M., & DeSalle, R. (2008). Mitochondrial DNA extraction and sequencing of formalin-fixed archival snake tissue. *Mitochondrial DNA*, *19*, 433–437.

Frost, D. R. (2020). *Amphibian species of the world: an online reference* Version 6. (pp. 1–23). Electronic Database. American Museum of Natural History. http://research.amnh.org/herpetology/amphibia/index.html

Funk, W. C., Caminer, M., & Ron, S. R. (2012). High levels of cryptic species diversity uncovered in Amazonian frogs. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1806–1814.

Giam, X., Scheffers, B. R., Sodhi, N. S., Wilcove, D. S., Ceballos, G., & Ehrlich, P. R. (2012). Reservoirs of richness: Least disturbed tropical forests are centres of undescribed species diversity. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 67–76.

Gotelli, N. J. (2004). A taxonomic wish-list for community ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*, 585–597.

Hadlich, H. L., Durgante, F. M., dos Santos, J., Higuchi, N., Chambers, J. Q., & Vicentini, A. (2018). Recognizing Amazonian tree species in the field using bark tissues spectra. *Forest Ecology and Management*, *427*, 296–304.

Hoogmoed, M. (2019). *Censo da Biodiversidade da Amazônia Brasileira*. Museu Goeldi -Plataforma digital. http://censo.museu-goeldi.br:8080/museugoeldi-web-1.2.0/paginas/especie_consultar.xhtml

Jaksch, K., Eschner, A., Rintelen, T. V., & Haring, E. (2016). DNA analysis of molluscs from a museum wet collection: A comparison of different extraction methods. *BMC Research Notes*, *9*, 348.

Jouquet, P., Capowiez, Y., Bottinelli, N., & Traoré, S. (2014). Potential of Near Infrared Refl ectance Spectroscopy (NIRS) for identifying termite species. *European Journal of Soil Biology*, *60*, 49–52.

Kaefer, I. L., Rojas, R. R., Ferrão, M., Farias, I. P., & Lima, A. P. (2019). A new species of Amazophrynella (Anura: Bufonidae) with two distinct advertisement calls. *Zootaxa*, *4577*, 316–334.

Kuhn, M. (2012). *The caret package*. Vienna, Austria: R Foundation for Statistical Computing. https://cran.r-project.org/package=caret

Lang, C., Almeida, D. R. A., & Costa, F. R. C. (2017). Discrimination of taxonomic identity at species, genus and family levels using Fourier Transformed Near-Infrared Spectroscopy (FT-NIR). *Forest Ecology and Management*, *406*, 219–227.

Lang, C., Costa, F. R. C., Camargo, J. L. C., Durgante, F. M., & Vicentini, A. (2015). Near infrared spectroscopy facilitates rapid identification of both young and mature amazonian tree species (D Gomory, Ed.). *PLoS One*, *10*: e0134521.

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, *405*, 243–253.

Meineke, E. K., Davies, T. J., Daru, B. H., & Davis, C. C. (2019). Biological collections for understanding biodiversity in the Anthropocene. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*, 20170386.

Miralles, A., Bruy, T., Wolcott, K., Scherz, M. D., Begerow, D., Beszteri, B., Bonkowski, M., Felden, J., Gemeinholzer, B., Glaw, F., Glöckner, F. O., Hawlitschek, O., Kostadinov, I., Nattkemper, T. W., Printzen, C., Renz, J., Rybalka, N., Stadler, M., Weibulat, T., … Vences, M. (2020). Repositories for taxonomic data: Where we are and what is missing. *Systematic Biology*, *69*(6), 1231–1253.

Moraes, L. J. C. L., Pavan, D., Barros, M. C., & Ribas, C. C. (2016). The combined influence of riverine barriers and flooding gradients on biogeographical patterns for amphibians and squamates in south-eastern Amazonia. *Journal of Biogeography*, *43*, 2113–2124.

Padial, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy review. *Frontiers in Zoology*, *7*(1), 1–14.

Pasquini, C. (2003). Near Infrared Spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, *14*, 198–219.

Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica Chimica Acta*, *1026*, 8–36.

Pavia, D. L., Lampman, G. M., Kriz, G. S., & Vyvyan, J. R. (2010). *Introdução à espectroscopia*. Cengage Learning.

Peloso, P. L. V., De, O. R. M., Sturaro, M. J., Rodrigues, M. T., Lima-Filho, G. R., Bitar, Y. O. C., Wheeler, W. C., & Aleixo, A. (2018). Phylogeny of map tree frogs, *Boana semilineata* Species Group, with a New Amazonian Species (Anura: Hylidae). *South American Journal of Herpetology*, *13*, 150.

Prata, E. M., Sass, C., Rodrigues, D. P., Domingos, F. M., Specht, C. D., Damasco, G., Ribas, C. C., Fine, P. V. A., & Vicentini, A. (2018). Towards integrative taxonomy in Neotropical botany: disentangling the Pagamea guianensis species complex (Rubiaceae). *Botanical Journal of the Linnean Society*, *188*(2), 213–231.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R.project.org/

Rigby, C. L., Wedding, B. B., Grauf, S., & Simpfendorfer, C. A. (2014). The utility of near infrared spectroscopy for age estimation of deepwater sharks. *Deep Sea Research Part I*, *94*, 184–194.

Rinnan, Å., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, *28*, 1201–1222.

Rodriguez-Fernandez, J. I., De Carvalho, C. J. B., Pasquini, C., Gomes De Lima, K. M., Moura, M. O., & Carbajal Arizaga, G. G. (2011). Barcoding without DNA? Species identification using near infrared spectroscopy. *Zootaxa*, *54*, 46–54.

Schander, C., & Halanych, K. M. (2005). DNA, PCR and formalinized animal tissue ? A short review and protocols. *Organisms Diversity and Evolution*, *3*, 195–205.

Simmons, J. E. (2002). Herpetological Collecting and Collections Management. *Herpetological Circular*.

Stevens, A., Ramirez-Lopez, L., Stevens, M. A., & Rcpp, L. (2020). *Package 'prospectr'*. R Package Version, (2).

Vance, C. K., Kouba, A. J., & Willard, S. T. (2014). Near infrared spectroscopy applications in amphibian ecology and conservation: Gender and species identification. *NIR News*, *25*, 10–15.

Vance, C. K., Tolleson, D. R., Kinoshita, K., Rodriguez, J., & Foley, W. J. (2016). Near infrared spectroscopy in wildlife and biodiversity. *Journal of Near Infrared Spectroscopy*, *24*, 1–25.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer.

Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). *klaR analyzing German business cycles. In Data analysis and decision support* (pp. 335–343). Berlin, Heidelberg: Springer.

Williams, B. K., & Titus, K. (1988). Assessment of sampling stability in ecological applications of discriminant analysis. *Ecology*, *69*, 1275–1285.

Workman, J. Jr, & Weyer, L. (2008). *Practical guide and spectral atlas for interpretive near-infrared spectroscopy*. CRC Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Figure S1. Understanding of the method.

Figure S2. Principal components analysis (PCA) of preprocessed data including samples of series 1.

Figure S3. Principal components analysis (PCA) of raw data including samples of series 1.

Figure S4. Principal components analysis (PCA) of preprocessed data including spectral-mean models (series 2).

Figure S5. Principal components analysis (PCA) of raw data including spectral-mean models (series 2).

Table S1. Complete list of selected variables for each body position of the individuals and the recognized functional group when known for mean models for back and belly with raw data.

Table S2. Complete list of selected variables for each body position of the individuals and the recognized functional group when known for mean models for back and belly with preprocessing.

Data S1. Plots of raw versus preprocessed spectra mean for species in all models tested.

How to cite this article: Torralvo K, Magnusson WE, Durgante F. Effectiveness of Fourier transform near-infrared spectroscopy spectra for species identification of anurans fixed in formaldehyde and conserved in alcohol: A new tool for integrative taxonomy. *J Zool Syst Evol Res*. 2020;00:1–14. https://doi.org/10.1111/jzs.12442

APPENDIX 1

Additional information specimen list.

Test	Species	Sample	Site	Colletion year
Dry	*Allobates tapajos*	INPA_H34402	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34403	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34404	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34405	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34406	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34407	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34408	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34409	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34410	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34411	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34412	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34413	PARNA da Amazonia, Pará	2007
Dry	*Allobates tapajos*	INPA_H34414	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32961	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32962	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32963	PARNA da Amazonia, Pará	2007

(Continues)

APPENDIX 1 (Continued)

Test	Species	Sample	Site	Colletion year
Dry	*Allobates magnussoni*	INPA_H32964	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32965	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32966	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32967	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32968	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32969	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32970	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32971	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32972	PARNA da Amazonia, Pará	2007
Dry	*Allobates magnussoni*	INPA_H32973	PARNA da Amazonia, Pará	2007
Dry	*Allobates caeruleodactylus*	INPA_H7229	Careiro	1995
Dry	*Allobates caeruleodactylus*	INPA_H7230	Careiro	1995
Dry	*Allobates caeruleodactylus*	INPA_H7231	Careiro	1995
Dry	*Allobates caeruleodactylus*	INPA_H7232	Careiro	1995
Dry	*Allobates caeruleodactylus*	INPA_H7234	Careiro	1995
Dry	*Allobates caeruleodactylus*	INPA_H7235	Careiro	1995
Dry	*Allobates caeruleodactylus*	INPA_H7236	Careiro	1995
Dry	*Allobates caeruleodactylus*	INPA_H7237	Careiro	1995
Dry	*Allobates caeruleodactylus*	INPA_H7238	Careiro	1995
Dry	*Amazophrynella bilinguis*	INPA_H39774	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39775	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39776	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39777	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39778	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39779	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39780	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39781	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39782	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39783	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39784	Faz. Taperinha, Santarem, PA	2011
Dry	*Amazophrynella bilinguis*	INPA_H39785	Faz. Taperinha, Santarem, PA	2011
Wet	*Adenomera andrea*	INPA-H036705	Resex Tapajós-Arapiuns	2013
Wet	*Adenomera andrea*	INPA-H036706	Resex Tapajós-Arapiuns	2013
Wet	*Adenomera andrea*	INPA-H036776	Resex Tapajós-Arapiuns	2013
Wet	*Adenomera andrea*	INPA-H036777	Resex Tapajós-Arapiuns	2013
Wet	*Adenomera andrea*	INPA-H036778	Resex Tapajós-Arapiuns	2013
Wet	*Adenomera andrea*	INPA-H036781	Resex Tapajós-Arapiuns	2013
Wet	*Adenomera andrea*	INPA-H036784	Resex Tapajós-Arapiuns	2013
Wet	*Adenomera andrea*	INPA-H036788	Resex Tapajós-Arapiuns	2013
Wet	*Adenomera andrea*	INPA-H036789	Resex Tapajós-Arapiuns	2013
Wet	*Adenomera andrea*	INPA-H036797	Resex Tapajós-Arapiuns	2013
Wet	*Adenomera andrea*	INPA-H037396	Presidente Figueiredo	2016
Wet	*Adenomera andrea*	INPA-H037397	Presidente Figueiredo	2016
Wet	*Adenomera andrea*	INPA-H037398	Presidente Figueiredo	2016
Wet	*Adenomera andrea*	INPA-H037399	Presidente Figueiredo	2016

(Continues)

APPENDIX 1 (Continued)

Test	Species	Sample	Site	Colletion year
Wet	*Adenomera andrea*	INPA-H037400	Presidente Figueiredo	2016
Wet	*Leptodactylus fuscus*	INPA-H001938	Lago Puraquequara	1993
Wet	*Leptodactylus fuscus*	INPA-H001939	Lago Puraquequara	1993
Wet	*Leptodactylus fuscus*	INPA-H001940	Lago Puraquequara	1993
Wet	*Leptodactylus fuscus*	INPA-H001941	Lago Puraquequara	1993
Wet	*Leptodactylus fuscus*	INPA-H001942	Lago Puraquequara	1993
Wet	*Adenomera hylaedatyla*	INPA_H002421	Rio Juruá, left bank	1992
Wet	*Adenomera hylaedatyla*	INPA_H005754	Rio Juruá, left bank	1992
Wet	*Adenomera hylaedatyla*	INPA_H005180	Rio Juruá, left bank	1992
Wet	*Adenomera hylaedatyla*	INPA_H005625	Rio Juruá, left bank	1992
Wet	*Adenomera hylaedatyla*	INPA_H019787	UFAM, Manaus	2007
Wet	*Adenomera hylaedatyla*	INPA_H019788	UFAM, Manaus	2007
Wet	*Adenomera hylaedatyla*	INPA_H019789	UFAM, Manaus	2007
Wet	*Adenomera hylaedatyla*	INPA_H019790	UFAM, Manaus	2007
Wet	*Adenomera hylaedatyla*	INPA_H019791	UFAM, Manaus	2007
Wet	*Adenomera hylaedatyla*	INPA_H034546	UFAM, Manaus	2007
Wet	*Adenomera hylaedatyla*	INPA_H034547	UFAM, Manaus	2007
Wet	*Amazophrynella minuta*	INPA-H12235	R.Madeira, Rondonia	2004
Wet	*Amazophrynella minuta*	INPA-H12236	R.Madeira, Rondonia	2004
Wet	*Amazophrynella minuta*	INPA-H12237	R.Madeira, Rondonia	2004
Wet	*Amazophrynella minuta*	INPA-H12248	R.Madeira, Rondonia	2004
Wet	*Amazophrynella minuta*	INPA-H12249	R.Madeira, Rondonia	2004
Wet	*Amazophrynella minuta*	INPA-H12324	R.Aripuanã,Amazonas	2004
Wet	*Amazophrynella minuta*	INPA-H12325	R.Aripuanã,Amazonas	2004
Wet	*Amazophrynella minuta*	INPA-H12327	R.Aripuanã,Amazonas	2004
Wet	*Amazophrynella minuta*	INPA-H12329	R.Aripuanã,Amazonas	2004
Wet and Dry	*Leptodactylus fuscus*	INPA-H007601	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007602	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007603	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007604	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007605	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007606	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007607	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007608	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007609	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007610	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007611	UHE Serra da mesa- Goias	1993
Wet and Dry	*Leptodactylus fuscus*	INPA-H007612	UHE Serra da mesa- Goias	1993
Wet and Dry	*Amazophrynella manaos*	INPA_H21864	ZF2, Região de Manaus	Unknown
Wet and Dry	*Amazophrynella manaos*	INPA_H21868	ZF2, Região de Manaus	Unknown
Wet and Dry	*Amazophrynella manaos*	INPA_H21869	ZF2, Região de Manaus	Unknown
Wet and Dry	*Amazophrynella manaos*	INPA_H21877	ZF2, Região de Manaus	Unknown
Wet and Dry	*Amazophrynella manaos*	INPA_H21884	ZF2, Região de Manaus	Unknown
Wet and Dry	*Amazophrynella manaos*	INPA_H21886	ZF2, Região de Manaus	Unknown
Wet and Dry	*Amazophrynella manaos*	INPA_H21891	ZF2, Região de Manaus	Unknown

(Continues)

APPENDIX 1 (Continued)

Test	Species	Sample	Site	Colletion year
Wet and Dry	*Amazophrynella manaos*	INPA_H21893	ZF2, Região de Manaus	Unknown
Wet and Dry	*Amazophrynella manaos*	INPA_H21897	ZF2, Região de Manaus	Unknown
Wet and Dry	*Amazophrynella manaos*	INPA_H21901	ZF2, Região de Manaus	Unknown
Wet and Dry	*Amazophrynella manaos*	INPA_H21903	ZF2, Região de Manaus	Unknown
Wet and Dry	*Amazophrynella manaos*	INPA_H21914	ZF2, Região de Manaus	Unknown
Wet and Dry	*Leptodactylus longirostris*	INPA-H019191	Parque Nacional do Viruá	2002
Wet and Dry	*Leptodactylus longirostris*	INPA-H019192	Parque Nacional do Viruá	2002
Wet and Dry	*Leptodactylus longirostris*	INPA-H019193	Parque Nacional do Viruá	2002
Wet and Dry	*Leptodactylus longirostris*	INPA-H019195	Parque Nacional do Viruá	2002
Wet and Dry	*Leptodactylus longirostris*	INPA-H019196	Parque Nacional do Viruá	2002
Wet and Dry	*Leptodactylus longirostris*	INPA-H001807	Reserva Ducke - Manaus	1994
Wet and Dry	*Leptodactylus longirostris*	INPA-H001808	Reserva Ducke - Manaus	1994
Wet and Dry	*Leptodactylus longirostris*	INPA-H001809	Reserva Ducke - Manaus	1994
Wet and Dry	*Leptodactylus longirostris*	INPA-H001810	Reserva Ducke - Manaus	1994
Wet and Dry	*Leptodactylus longirostris*	INPA-H001811	Reserva Ducke - Manaus	1994
Wet and Dry	*Phyllomedusa bicolor*	INPA-H2234	R.Jurua, Amazonas	1991
Wet and Dry	*Phyllomedusa bicolor*	INPA-H2964	R.Jurua, Amazonas	1991
Wet and Dry	*Phyllomedusa bicolor*	INPA-H3574	R.Jurua, Amazonas	1991
Wet and Dry	*Phyllomedusa bicolor*	INPA-H3675	R.Jurua, Amazonas	1991
Wet and Dry	*Phyllomedusa bicolor*	INPA-H8437	Res.Gavião, Amazonas	1995
Wet and Dry	*Phyllomedusa bicolor*	INPA-H8438	Res.Gavião, Amazonas	1995
Wet and Dry	*Phyllomedusa bicolor*	INPA-H8439	Res.Gavião, Amazonas	1995
Wet and Dry	*Phyllomedusa bicolor*	INPA-H8440	Res.Gavião, Amazonas	1995
Wet and Dry	*Phyllomedusa bicolor*	INPA-H8441	Res.Gavião, Amazonas	1995
Wet and Dry	*Callimedusa tomopterna*	INPA-H6802	Anavilhanas, Amazonas	1996
Wet and Dry	*Callimedusa tomopterna*	INPA-H6815	Anavilhanas, Amazonas	1996
Wet and Dry	*Callimedusa tomopterna*	INPA-H6817	Anavilhanas, Amazonas	1996
Wet and Dry	*Callimedusa tomopterna*	INPA-H6819	Anavilhanas, Amazonas	1996
Wet and Dry	*Callimedusa tomopterna*	INPA-H30959	M.Taboca, Amazonas	2012
Wet and Dry	*Callimedusa tomopterna*	INPA-H30960	M.Taboca, Amazonas	2012
Wet and Dry	*Callimedusa tomopterna*	INPA-H30961	M.Taboca, Amazonas	2012
Wet and Dry	*Callimedusa tomopterna*	INPA-H30962	M.Taboca, Amazonas	2012
Wet and Dry	*Callimedusa tomopterna*	INPA-H30963	M.Taboca, Amazonas	2012
Wet and Dry	*Callimedusa tomopterna*	INPA-H30964	M.Taboca, Amazonas	2012