Discrimination of termite species using Near-Infrared Spectroscopy (NIRS)

Renato Almeida de Azevedo¹,*, José Wellington de Morais¹, Carla Lang¹, Cristian de Sales Dambrosb

¹ Coordenação de Biodiversidade-COBIO, Instituto Nacional de Pesquisas da Amazônia-INPA, Av. André Araújo, 2936, Petrópolis, 69083-000, Manaus, AM, Brazil
² Universidade Federal de Santa Maria, Departamento de Ecologia e Evolução, Av. Roraima, 1000-7, Camobi, 97105-900, Santa Maria, RS, Brazil

ABSTRACT

The classification of most taxa is based exclusively on morphological characters which have limited capacity to discriminate species with phenotypic plasticity and to detect the existence of cryptic species. These limitations can be reduced by the use of recent techniques that quantify molecular attributes in a sample. Near-Infrared Spectroscopy (NIRS) is a non-destructive method that has been used as an efficient tool to discriminate species from several plant and animal groups in controlled laboratory conditions. We tested the efficacy of this method as an alternative to morphology to discriminate termite castes, species, genera, and families. Seven species were selected: Cylindrotermes flangius; Cylindrotermes parvignathus; Cornitermes pugnax; Cornitermes ovatus; Heterotermes tenuis; Heterotermes crinitus and Coptotermes testaceus. Discriminant models were created with the separation of 70% of the samples for training (creation of the model) and 30% for testing. We found a high level of accuracy for the discrimination of families, genera, species and castes, ranging from 76% to 100% of accuracy (mean of 90%). This high level of accuracy, using the soldier and worker castes, indicates that Near-Infrared Spectroscopy serves as a reliable alternative to identify termite species.

1. Introduction

Species classification is traditionally based on the differentiation of species using external morphological characters. The use of morphological characters is limited as the method usually does not take into account phenotypic plasticity or the existence of cryptic taxa [1]. In addition, the classification of many groups requires a high level of taxonomic expertise [2]. Recent advances in molecular and spectroscopic techniques have allowed taxonomists to expand their toolbox and to include multiple and complementary perspectives to identify and classify species [3]. However, most of these new techniques have not been automated as tools for species discrimination.

The Near-Infrared Spectroscopy (NIRS) has been a reliable alternative to discriminate plants [4–6], insects [7–9], and bacteria [10–14], including cryptic species with no clear morphological differentiation [15]. NIRS detects signs of organic, chemical, and structural compounds in a sample [16] and is able to quantify molecular components in biological material [17]. NIRS is a very rapid method (approximately a full spectral reading per second), does not generate solid, liquid or gaseous wastes, being a clean technology (environmentally or ecologically viable), requires only small amounts of sample, can be implemented in situ, and is not invasive or destructive [18,19]. In addition, NIRS can be used with fresh or dried samples, making it more advantageous than many other classification techniques, which are usually destructive and costlier, such as DNA barcoding, NIRS has been shown to be highly accurate for termite classification [8,20]. However, the single previous study to classify termite species using the technique used macerated samples, which is not ideal for rare specimens or type material. Maceration may also influence spectral readings because food and particles present in the termite gut may directly interfere with NIRS readings. The gut represents a large proportion of the termite body and constitutes a microenvironment with specific environmental and microbial properties that could be identified in macerated samples instead of the individuals under examination [21,22]. To date, the ability of NIRS to discriminate termites has never been evaluated using intact samples, to compare several species, genera and families simultaneously, and using samples obtained from natural field sampling conditions (in contrast to laboratory cultivated colonies).

Termite identification and classification is mainly based on the use of morphological characters of soldier caste. Therefore, if the soldier is not found or species of soldierless termites are found, the taxonomic classification is extremely difficult [23], often requiring comparison using internal morphology or molecular data [24]. We tested the efficiency of Near Infrared Spectroscopy in...
discriminating 381 termite colonies from seven species sampled in central Amazonia, Brazil. We also tested for the accuracy of the method to discriminate families and genera, and the ability of this technique to discriminate species using only the soldier or worker castes.

2. Material and methods

2.1. Termite sampling

Termites were previously sampled in Reserve Ducke (3°05′0 S, 60°00′0 W), Manaus, Brazil [25]. The sampling area is covered by a moist *terra-firme* forest that is not subjected to periodic flooding [2]. The climate is classified as tropical humid, with humidity ranging from 75 to 86% and annual precipitation of 1,750 to 2,500 mm concentrated from November to May [26]. The topography forms a mosaic of plateaus, slopes and bottomlands [27]. The altitude varies from 40 to 140 m a.s.l., and soil texture in the sampling area was clayey in the ridges and sandy in valleys [28].

The samples were collected in soil, leaf litter, rotting logs, and tree and shrub roots. The selection of *Cylindrotermes flavigatus* Mathews, *Cylindrotermes parvignathus* Emerson in Snyder, *Cornitermes pugnax* Emerson, *Cornitermes ovatus* Emerson, *Heterotermes tenuis* (Hagen), *Heterotermes crinitus* (Emerson), *Coptotermes testaceus* (Linnaeus), was based on the following criteria: 1) species are represented by multiple colonies along edaphic gradients; 2) selected genera belong to more than one family for family level comparisons; 3) species have distinct morphological characters in the soldier caste and are easy to classify using morphological characters when the soldier caste is present; 4) species have a well-established taxonomy (all genera have been reviewed in the literature after description and all species were included in previous revisions). All samples were identified by specialists and by comparison with museum collections at the Federal University of Rio Grande do Norte and the National Institute of Amazonian Research (INPA), Brazil. Therefore, failure in NIRS to discriminate termites cannot be attributed to incorrect taxonomic classification.

2.2. Spectral readings

We recorded the spectral absorption in 402 samples of termite colonies between 4,000 and 10,000 cm⁻¹ (2.5 × 10⁶ and 10 × 10⁶ nm) in the near infrared spectrum. These readings were conducted using the Thermo Nicolet spectrophotometer, Antaris II FT-NIR Method Development System (MDS) in the wood chemistry laboratory of the National Institute of Amazonian Research, INPA, Manaus, Brazil. The equipment employed use diffuse reflection. When the light was directed to the surface, the infrared radiation interacts with the surface, alternating and reflecting the light (Fig. S1 in Supporting Material S1).

Six to ten colonies were selected for each termite species (Table 1). We selected three soldiers and three workers from each colony and performed a spectral reading per individual. All specimens used in this study were intact, with head capsule, thorax, abdomen, and appendices complete. Termite samples were stored in 70% EtOH which can be identified in spectral readings and interfere with spectral sampling. To avoid interference, the specimens were placed to dry in a Petri dish covered with a paper towel until the individual was completely dry (2–4 min). Although alcohol could be identified a posteriori, was preferred to standardize the spectral reading to maintain the analyses as simple as possible and to avoid removing parts of the spectra that could be informative. For plants, previous studies have not found differences between fresh and dry samples [5].

To standardize spectral sampling, all specimens were placed in the same position, and the reading was taken on the dorsal position (Fig. S1). In addition, to avoid contamination between readings, individuals were removed from the Petri dish one at a time and separated to prevent the repeated use of the same individual. It was used a cup specifically designed for the spectrophotometer to avoid light dispersion (Fig. S1). In each reading, spectral readings were plotted to identify possible reading errors caused by mispositioning the individual, by the presence of alcohol remnants, or soil material not visible in the sample (Fig. S2). Based on visual check, a total of 21 spectral readings were clearly misreadings not recognizable as a biological sample (eg, flat reading – same reading in all wavelengths; characteristic peaks missing in the longer wavelengths), and were excluded from the data. The final database had 381 spectra used for analysis. Multiple readings per sample can be performed and averaged to increase the precision of spectral measurements [9]. However, it was opted to increase the number of samples obtained instead of the precision in individual samples. To keep the analyses as simple as possible, it was also not used any noise removal technique. Considering the high quality of the results, it is likely that these techniques would cause only minor improvements.

The NIRs can analyze any type of molecule [18], being related to the nature of the molecular bonds, which in turn are defined by the bonds between atoms and or groups of atoms (functional groups) that form the sample. In this way, it is possible to identify all the molecular components that the sample has and to separate the parts of the spectrum that are most important to discriminate the sample.

2.3. Data analysis

To test the ability of NIR data to discriminate spectra from individual specimens at the family, gender, species, and caste level, we created a Linear Discriminant (LDA) model. LDA was used to determine the potential of FT-NIR in distinguishing species correctly. The dependent variables were the groups compounded by predefined taxa and independent variables were the absorbance value in each wavelength read. This method has been used successfully to discriminate species [5,6,29,30]. Three distinct discriminant models were created using 1)
all individuals present, 2) only soldiers, and 3) only workers. In all models, 70% of the spectra were used for model fitting (training) and the remaining 30% of the spectra were used for testing. Since distinct combinations can be generated by selecting 70% of the spectra, the selection for training and testing was performed randomly 999 times. In each randomization, random (and different) combinations of spectra were selected to represent the training and testing groups, and the number of correct and incorrect identifications was quantified using the randomly selected spectra. As in most analyses using training and testing data in discriminant models [4–6], the average of correct identifications in the 999 randomizations was used to calculate the accuracy of the model (percentage of correct identifications).

All analyzes were performed in the R program [31] using the MASS package [32] and functions specifically created for this paper. A documented script explaining all analyzes performed, the R functions used, and the original spectral readings are available as Supporting Information (Supplementary Material S2).

3. Results

The discriminant model using all castes obtained a success rate that ranged from 77%, of correct identifications for Cylindrotermes parvignathus, to 100% for Heterotermes tenuis. When using only the soldier caste, this margin remained similar, with 76% of correct identifications for Cornitermes pugnax and 100% for Heterotermes tenuis. Using only the worker caste, we obtained 78% correct identifications for Cylindrotermes parvignathus and 100% for Heterotermes tenuis. In relation to the higher taxonomic levels, there was a higher percentage of correct identifications for genera (84–99%) and families (92–99%) than for species (Table 2).

Regarding the generated models, when used the spectra of all individuals present, most of the errors were found in the same genus, with the exception of 10% of Coptotermes identified as Cylindrotermes and 10% of Cap. testaceus identified as Cyl. flangiatus (Fig. 1A). The model, using only soldier, was the single one that presented an error at the family level, with 10% of Rhinotermitidae identified as Termididae; at the genus level, 10% of Heterotermes and Coptotermes were identified as Cylindrotermes; and at the species level, 10% of H. crinitus was...
identified as *Cyl. parvignathus* and 20% of *Cop. testaceus* was wrongly identified, being 10% as *Het. crinitus* and 10% as *Cyl. flangiatus* (Fig. 1B). With the model generated only with workers, at the genus level, 10% of *Coptotermes* and *Cornitermes* were identified as *Cylindrotermes*, and at the species level, 10% of *Cop. testaceus* and *Cor. pugnax* were identified as *Cyl. flangiatus*, the other errors were within the same genus (Fig. 1C).

4. Discussion

Our results indicate that Near-Infrared Spectroscopy (NIRS) can be very effective as a technique to discriminate species and termite castes for the majority of species – more than 90% of the samples were correctly identified (Table 2). This is the first study to demonstrate that termite classification using only termite workers is reliable and comparable to the classification using soldiers. Soldiers and workers likely have similar sets of cuticular hydrocarbons for all castes [33]. Therefore, identification of species based on a given caste (e.g. worker) is possible even for those species in which only other castes (e.g. soldier) are known.

Although the method was highly accurate for most termite species, there were important differences in method accuracy. For example, 100% of samples of *Heterotermes tenuis* were correctly identified, whereas only 92% of samples of *Coptotermes testaceus* were correctly identified, both species belonging to the Rhinotermitidae family. The difficulty to distinguish species was especially important for termites living in the soil or with close contact to the soil. These results might suggest that spectral readings were affected by remnants of the environment. Near-Infrared Spectroscopy quantifies molecular phenotypes, which are the mean of the vibration modes of all the molecules present in the specimen [9]. However, the method can be affected by any component present in the sample, such as soil or alcohol [18]. Although this might limit the use of NIR spectra to identify new samples, especially those obtained environments not included in model training, future improvements in spectral discrimination could increase model performance.

In this study, the spectrum was used in its raw state considering all the absorbance values from each wavelength. However, it also can be used only the portion in which the wavelength better separate species. Despite species delimitation being a topic of permanent discussion [34], the NIR has been an effective technique to discriminate well-known species [4–6].

In addition to the identification of discriminant wavelengths, many improvements of the technique can still be made by the use of other statistical methods with greater predictive power [35]. We used simple analyses and samples obtained from normal field conditions to test the efficiency of the method. Therefore, we tried to replicate the conditions in which termite discrimination using the technique would be more useful and make the methods as simple as possible to future users. The results presented are highly encouraging and demonstrate that this technique has a high potential to aid in species discrimination.

Acknowledgements

We thank Dr. Niro Hijuchi, Dr. Elizabeth Franklin and Dr. Flávia Durgante by all the support. Also, we thank Rafael Sobral for kindly reviewing the grammar of the manuscript. We acknowledge the Conselho Nacional de Desenvolvimento Científico e Tecnológico, the Instituto Nacional de Pesquisas da Amazônia for research support, and the Fundação de Amparo à Pesquisas do Estado do Amazonas (FAPEAM) for the Ph.D. scholarship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejsobi.2019.04.002.

References


