Contents lists available at ScienceDirect

European Journal of Soil Biology

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

Renato Almeida de Azevedo[a,*], José Wellington de Morais[a], Carla Lang[a], Cristian de Sales Dambros[b]

[a] Coordenação de Biodiversidade-COBIO, Instituto Nacional de Pesquisas da Amazônia-INPA, Av. André Araújo, 2936, Petrópolis, 69083-000, Manaus, AM, Brazil
[b] 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 flangiatus*; *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 spectrometric 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 tested as automated 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 flangiatus* 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} (2.5×10^6 and 10×10^6 nm) in the near infrared spectrum. These readings were conducted using the Thermo Nicollet 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

Table 1
Species with their respective number of colonies and spectra collected from their respective castes.

Species		Caste	
	Colonies	Soldier	Worker
Cylindrotermes flangiatus Mathews	10	30	30
Cylindrotermes parvignathus Emerson in Snyder	10	30	30
Cornitermes pugnax Emerson	10	30	27
Cornitermes ovatus Emerson	10	30	30
Heterotermes tenuis (Hagen)	10	30	30
Heterotermes crinitus (Emerson)	6	18	18
Coptotermes testaceus (Linnaeus)	10	30	18
Total	66	198	183

Table 2
Correctness percentage of the discriminant models using all the spectra (Complete), percentage using only the spectra of the caste soldier, and percentage using only the spectra of the caste worker in relation to taxonomic categories.

Categorie		% of correctness		
	Taxon	Complete	Soldier	Worker
Family	Rhinotermitidae	0.97	0.92	0.96
	Termitidae	0.99	0.98	0.99
Genus	*Coptotermes*	0.92	0.84	0.93
	Cornitermes	0.96	0.99	0.91
	Cylindrotermes	0.99	0.99	0.99
	Heterotermes	0.98	0.91	0.97
Species	*Cylindrotermes flangiatus*	0.85	0.89	0.87
	Cornitermes ovatus	0.93	0.93	0.93
	Cornitermes pugnax	0.88	0.76	0.81
	Cylindrotermes parvignathus	0.77	0.84	0.78
	Coptotermes testaceus	0.91	0.83	0.92
	Heterotermes crinitus	0.99	0.93	0.99
	Heterotermes tenuis	1	1	1

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 distinguish species correctly. The dependent variables were the groups compound 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)

Fig. 1. Matrices with the results from discriminant models. The taxonomic categories are the columns and taxa names are the lines. A – using all the spectra; B – using the spectra of the caste soldier; C – using the spectra of the caste worker. Red color represents families, blue represents genera, green represents species and yellow represents castes. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

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 *Cop. 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 Termitidae; at the genus level, 10% of Heterotermes and Coptotermes were identified as Cylindrotermes; and at the species level, 10% of *Het. 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 *Corniterms* 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

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

[1] A.C.S. Gomes, A. Andrade, J.S. Barreto-Silva, T. Brenes-Arguedas, D.C. López, C.C. Freitas, C. Lang, A.A. Oliveira, A.J. Pérez, R. Perez, J.B. Silva, A.M.F. Silveira, M.C. Vaz, J. Vendrami, A. Vicentini, Local plant species delimitation in a highly diverse Amazonian forest: do we all see the same species? J. Veg. Sci. 24 (2013) 70–79 https://doi.org/10.1111/j.1654-1103.2012.01441.x.

[2] M.J.G. Hopkins, Flora da Reserva Ducke, Amazonas, Brasil, Rodriguesia 56 (2005) 9–25 https://doi.org/10.1590/2175-78602005568602.

[3] B. Dayrat, Towards integrative taxonomy, Biol. J. Linn. Soc. 85 (2005) 407–415 https://doi.org/10.1111/j.1095-8312.2005.00503.x.

[4] F.M. Durgante, N. Higuchi, A. Almeida, A. Vicentini, Species spectral signature: discriminating closely related plant species in the Amazon with Near-Infrared Leaf-Spectroscopy, For. Ecol. Manag. 291 (2013) 240–248 https://doi.org/10.1016/j.foreco.2012.10.045.

[5] C. Lang, F.R.C. Costa, J.L.C. Camargo, F.M. Durgante, A. Vicentini, Near infrared spectroscopy facilitates rapid identification of both young and mature Amazonian tree, PLoS One 10 (2015) 1–15 https://doi.org/10.1371/journal.pone.0134521.

[6] C. Lang, D.R.A. Almeida, F.R.C. Costa, Discrimination of taxonomic identity at species, genus and family levels using Fourier Transformed Near-Infrared Spectroscopy (FT-NIR), For. Ecol. Manag. 406 (2017) 219–227 https://doi.org/10.1016/j.foreco.2017.09.003.

[7] F.E. Dowell, J.E. Throne, D. Wang, J.E. Baker, Identifying stored-grain insects using near-infrared spectroscopy, JEE (J. Econ. Entomol.) 92 (1999) 165–169 https://doi.org/10.1093/jee/92.1.165.

[8] B.T. Aldrich, E.B. Maghirang, F.E. Dowell, S. Kambhampati, Identification of termite species and subspecies of the genus *Zootermopsis* using near-infrared reflectance spectroscopy, J. Insect Sci. 7 (2007) 1–7 https://doi.org/10.1673/031.007.1801.

[9] J.I.R. Fernández, C.J.B. Carvalho, C. Pasquini, K.M.G. Lima, M.O. Moura, G.G.C. Arízaga, Barcoding without DNA? Species identification using near infrared spectroscopy, Zootaxa 2933 (2011) 46–54.

[10] D. Naumann, Infrared spectroscopy in microbiology, in: R.A. Meyers (Ed.), Encyclopedia of Analytical Chemistry, John Wiley & Sons, Ltd., Chichester, 2000, pp. 102–131.

[11] H. Oberreuter, H. Seiler, S. Scherer, Identification of coryneform bacteria and related taxa by Fourier-transform infrare d (FT-IR) spectroscopy, Int. J. Syst. Evol. Microbiol. 52 (2002) 91–100, https://doi.org/10.1099/00207713-52-1-91.

[12] E. Stackebrandt, W. Frederiksen, G.M. Garrity, P.A.D. Grimont, P. Kämpfer, M.C.J. Maiden, X. Nesme, R. Rosselló-Mora, J. Swings, H.G. Truper, L. Vauterin, A.C. Ward, W.B. Whitman, Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology, Int. J. Syst. Evol. Microbiol. 52 (2002) 1043–1047 https://dx.doi.org/10.1099/00207713-52-3-1043.

[13] S. Lai, R. Goodacre, L.N. Manchester, Whole-organism fingerprinting of the Genus Carnobacterium using fourier transform infrared spectroscopy (FT-IR), Syst. Appl. Microbiol. 27 (2004) 186–191 https://doi.org/10.1078/072320204322881808.

[14] C.L. Winder, E. Carr, R. Goodacre, R. Seviour, The rapid identification of Acinetobacter species using Fourier transform infrared spectroscopy, J. Appl. Microbiol. 96 (2004) 328–339 https://doi.org/10.1046/j.1365-2672.2003.02154.x.

[15] K.R. Girija, Ch Sasikala, C.H.V. Ramana, C. Spröer, S. Takaichi, V. Thiel, J.F. Imhoff, *Rhodobacter johrii* sp. nov., an endospore-producing cryptic species isolated from semi-arid tropical soils, Int. J. Syst. Evol. Microbiol. 60 (2010) 2099–2107 https://doi.org/10.1099/ijs.0.011718-0.

[16] S. Tsuchikawa, A review of recent near infrared research for wood and paper, Appl. Spectrosc. Rev. 42 (2007) 43–71 https://doi.org/10.1080/05704920601036707.

[17] A. Lima, J. Bakker, Near-infrared spectroscopy for monitoring peripheral tissue perfusion in critically ill patients, Revista Brasileira de Terapia Intensiva 23 (2011) 341–351 https://doi.org/10.1590/S0103-507X2011000300013.

[18] C. Pasquini, Near infrared spectroscopy: fundamentals practical aspects and analytical applications, J. Braz. Chem. Soc. 14 (2003) 138–219 https://doi.org/10.1590/S0103-50532003000200006.

[19] L. Munck, A new holistic exploratory approach to systems biology by near infrared spectroscopy evaluated by chemometrics and data inspection, J. Chemom. 21 (2007) 406–426 https://doi.org/10.1002/cem.1079.

[20] P. Jouquet, Y. Capowiez, N. Bottinelli, S. Traoré, Potential of near infrared reflectance spectroscopy (NIRS) for identifying termite species, Eur. J. Soil Biol. 60 (2014) 49–52 https://doi.org/10.1016/j.ejsobi.2013.11.004.

[21] A. Brune, Termite guts: the world's smallest bioreactors, Trends Biotechnol. 16 (1998) 16–21 https://doi.org/10.1016/S0167-7799(97)01151-7.

[22] A. Brune, M. Friedrich, Microecology of the termite gut: structure and function on a microscale, Curr. Opin. Microbiol. 3 (2000) 263–269 https://doi.org/10.1016/S1369-5274(00)00087-4.

[23] R. Constantino, Padrões de diversidade e endemismo de térmitas no bioma cerrado, in: A.O. Scariot, J.C.S. Silva, J.M. Felfili (Eds.), Biodiversidade, Ecologia e Conservação do Cerrado. Ministério do Meio Ambiente, Brasília, 2005, pp. 319–333.

[24] T. Bourguignon, R.H. Scheffrahn, J. Křeček, Z.T. Nagy, G. Sonet, Y. Roisin, Towards a revision of the Neotropical soldierless termites (Isoptera: Termitidae): redescription of the genus *Anoplotermes* and description of *Longustitermes* gen. nov, Invertebr. Syst. 24 (2010) 357–370 https://doi.org/10.1071/IS10012.

[25] C.S. Dambros, J.W. Morais, A. Vasconcellos, J.L.P. Souza, E. Franklin, N.J. Gotelli, Association of ant predators and edaphic conditions with termite diversity in an amazonian rain forest, Biotropica 0 (2016) 1–9 https://doi.org/10.1111/btp.12270.

[26] A.O. Marques-Filho, M.N.G. Ribeiro, J.M. Santos, Estudos climatológicos da Reserva Florestal Ducke, Manaus, AM. IV – precipitação, Acta Amazonica 4 (1981) 759–768 https://doi.org/10.1590/1809-43921981114759.

[27] J.E.L.S. Ribeiro, M.J.G. Hopkins, A. Vicentini, C.A. Sothers, M.A.S. Costa, J.M. Brito, M.A.D. Souza, L.H.P. Martins, L.G. Lohmann, P.A.C.L. Assunção, E.C. Pereira, C.F. Silva, M.R. Mesquita, L. Procópio, Flora da Reserva Ducke - guia de identificação das plantas vasculares de uma floresta de terra firme na Amazônia Central, INPA/DFID, Manaus, Brasil, 1999.

[28] A. Chauvel, Y. Lucas, R. Boulet, On the genesis of the soil mantle of the region of Manaus, central Amazonia, Brazil, Experientia 43 (1987) 234–241.

[29] J.B. Féret, G. Asner, Spectroscopic classification of tropical forest species using radiative transfer modeling, Rem. Sens. Environ. 115 (2011) 2415–2422 https://doi.org/10.1016/j.rse.2011.05.004.

[30] A. Carballo-Meilan, A. Goodman, M. Baron, J. Gonzalez Rodriguez, An specific case in the classification of woods by FTIR and chemometrics: discrimination of Fagales from Malphigiales, Cellulose 21 (2014) 261–273, https://doi.org/10.1007/s10570-013-0093-2.

[31] R. R Core Team, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2018.

[32] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, fourth ed., Springer, New York, 2002.

[33] S.G. Klochkov, V.I. Kozlovskii, N.V. Belyaeva, Caste and population specificity of termite cuticle hydrocarbons, Chem. Nat. Comp. 41 (2005) 1–6, https://doi.org/10.1007/s10600-005-0060-8.

[34] J.J. Wiens, Species delimitation: new approaches for discovering diversity, Syst. Biol. 56 (2008) 875–878 https://doi.org/10.1080/10635150701748506.

[35] T. Naes, K.I. Hildrum, Comparison of multivariate calibration and discriminant analysis in evaluating NIR spectroscopy for determination of meat tenderness, Appl. Spectrosc. 51 (1997) 350–357 https://doi.org/10.1366%2F0003702971940198.