SciELO Preprints

Publication status: Not informed by the submitting author

Spectral signatures as evidence to test hypotheses in plant species complexes

Niksoney Azevedo Mendonça, Marise Helen Vale Oliveira, Thaís Elias Almeida

https://doi.org/10.1590/SciELOPreprints.10972

Submitted on: 2024-12-20
Posted on: 2025-01-13 (version 1)
(YYYY-MM-DD)

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Spectral signatures as evidence to test hypotheses in plant species complexes

Niksoney Azevedo Mendonça[1*] Marise Helen Vale de Oliveira[2] and Thaís Elias Almeida[1]

[1] Departamento de Botânica, Programa de Pós-graduação em Biologia Vegetal, Centro de Biociências, Universidade Federal de Pernambuco, Recife - PE, Av. Professor Moraes Rego, 1235, 50670-420, Brasil. NAM: https://orcid.org/0000-0002-5336-011X, TEA: https://orcid.org/0000-0002-1611-1333

[2] Programa de Pós-graduação em Botânica, Instituto Nacional de Pesquisas da Amazônia, Manaus - AM, Av. André Araújo, 2936, 69067-375, Brasil. MHVO: https://orcid.org/0000-0002-1161-7378

*Corresponding author: niksoney.azevedo@ufpe.br

Short title: Spectral signatures in plant species complexes

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Abstract

Traditional methods for species definition, based on macromorphological characteristics, face limitations due to the high phenotypic plasticity observed in plants, which makes it challenging to accurately identify species complexes. Fourier transform near-infrared spectroscopy emerges as a promising, non-destructive technique for analyzing plant material, enabling the distinction of species. This study focuses on the Scaly clade of the fern genus *Microgramma*, characterized by complex taxonomic boundaries and morphological variations. A total of 94 samples from eight species, including fertile and sterile leaves of dimorphic and monomorphic species, were evaluated to test the effectiveness of FT-NIR in distinguishing these lineages. The average identification accuracy ranged from 86% to 91%, depending on the models and validation employed. Species with better-defined morphological characteristics, such as *Microgramma percussa*, achieved an accuracy of 100%. Conversely, species with higher morphological overlap showed lower accuracy, which may be associated with hybridization, introgression, or cryptic variation. Dimorphic species, with morphologically distinct fertile and sterile fronds, exhibited higher intraspecific spectral variation compared to monomorphic species, which may explain their lower accuracy rates. Fertile fronds, in some cases, provided more informative data, possibly due to the presence of sori increasing the complexity of spectra. This study highlights the potential of FT-NIR as a complementary tool in plant systematics. However, further research is needed to understand the influence of processes such as hybridization and features such as the indumentum on spectral readings. Overall, FT-NIR presents itself as a promising method to elucidate species limits in ferns and improve knowledge about their diversity.

Keywords: ferns, FT-NIR spectroscopy, *Microgramma*, predictive models, species limits, systematics.

1 Introduction

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

The continuous evolution of lineages drives the necessity to develop new approaches to explore biodiversity and define taxonomic boundaries (Wiens, 2007; Hörandl, 2022). Traditional methods relying solely on macromorphological features constitute emerging challenges to our understanding of biodiversity (Dayrat, 2005). Species are often defined one-dimensionally, which presents significant challenges (De Queiroz, 2007). They are generally not constructed or considered testable hypotheses and are often described in ways that impede their validation (De Queiroz, 2007; Sites & Marshall, 2003; Wiens, 2007). Additionally, plants tend to exhibit greater phenotypic plasticity compared to animals (Huey et al., 2002; Borges, 2008; Palacio-López et al., 2015), making it even more challenging to understand the phenotypic complexity of this group.

An accurate and multidimensional definition of species boundaries is crucial for understanding the diversity of living organisms, allowing us to determine whether their definition correspond or not to a single lineage (Dayrat, 2005; Will et al., 2005). This is particularly important since species complexes represent a significant part of biodiversity (Pinheiro et al., 2018). Despite significant genetic divergences, these complexes are characterized by low levels of morphological disparity between lineages, making their identification more difficult (Struck et al., 2018). Integrating data from different sources can overcome the limitations of traditional methods (Yeates et al., 2011; Edwards & Knowles, 2014; Cheng et al., 2021) based mainly on macromorphological characters that are not always diagnostic and can sometimes be homoplastic (Quattrini et al., 2019). To address this, approaches incorporating data from different levels of biological organization (Dayrat, 2005) promote complementarity between fields of study and encourage collaboration among specialists. Emerging methods can help us understand taxonomically recalcitrant groups, advancing our knowledge of biological diversity and the evolutionary processes that shape it (Rouhan & Gaudeul, 2021; Sandall et al., 2023).

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

One method in evidence is Fourier-transform near-infrared spectroscopy (FT-NIR) (Pasquini, 2018). This technique is promising due to its speed, accessibility, non-destructive nature, and lack of sample pre-treatment (Rodríguez-Fernández et al., 2011). The fundamental principle of the method involves exposing fragments of biological material (e.g., a dry leaf) to infrared radiation (Durgante et al., 2013; Paiva et al., 2021). The spectra generated by the material represent the energy absorbed by the C - H, C - N, and C - O bonds based on the amount of light absorbed (Pasquini, 2003, 2018). This process generates complex spectra that allow the sample's chemical and physical structure to be analyzed (Workman & Weyer, 2007; Durgante et al., 2013). Spectroscopy is the tool of choice in a wide range of studies, from those focused on understanding ontogeny (Fernandes et al., 2020) to plant stress (Zahir et al., 2022) and nutrient analysis of leaf tissues (Prananto et al., 2020), for example. It is also used to delimit and discriminate species, genera, and families (Xu et al., 2009; Prata et al., 2018), as well as for the recognition of new plant species (Vasconcelos et al., 2021; Gaem et al., 2020). In the context of spectral studies on ferns, only one study has been conducted specifically on this group, aiming to explore the method's efficiency (Paiva et al., 2021).

Ferns, the second most diverse group of vascular land plants (Nitta et al., 2022), are widely distributed around the world (Suissa et al., 2021). Their greatest diversity is found in tropical regions (Suissa et al., 2021). These plants, which lack flowers and fruits, have a life cycle characterized by alternating generations, including both sporophyte and gametophytic phases (Haufler et al., 2016). Since most lineages show uncommon cryptic variations (Kinosian et al., 2020; Ekrt et al., 2022; Wei et al., 2022; Yi et al., 2023), frequent hybridization events (Bloesch et al., 2022; Luo et al., 2024; Mendez-Reneau et al., 2024), and polyploidy (Fujiwara et al., 2023; Heslop-Harrison et al., 2023), it has proven challenging to delimit species based solely on morphological analyses. Although polyploidy is a crucial factor in speciation (Wood et al.,

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

2009; Alix et al., 2017), hybridization is equally significant, contributing to the emergence of new lineages and adding complexity to evolutionary processes (Sigel, 2016).

Our study focuses on the Scaly clade of the genus *Microgramma* C.Presl (Polypodiaceae), whose main synapomorphy is the presence of subulate and/or round scales on both sides of the leaf surfaces (sensu Almeida et al., 2021). The clade, with ca. 26 mya, was recovered in a recent study as the oldest lineage sister to all other clades within the genus (Almeida et al., 2021). With a distribution that covers the entire neotropical region, it includes both species with distinct morphological traits and species with high morphological overlaps, which hinder taxonomic identification. Some species are widely distributed, such as *M. dictyophylla* (Kunze ex Mett.) de la Sota, *M. nana* (Liebm.) T.E.Almeida, *M. percussa* (Cav.) de la Sota, *M. reptans* (Cav.) A.R.Sm., and *M. tobagensis* (C.Chr.) C.D.Adams & Baksh.- Com., while others have restricted distributions, such as *M. piloselloides* (L.) Copel. (Central America and the Caribbean), or are microendemic, such as *M. latevagans* (Maxon & C.Chr.) Lellinger (Bolivia and Peru) and *M. tecta* (Kaulf.) Alston (Brazilian Atlantic Forest) (Almeida et al., 2021). The morphological disparity in the clade is also significant, with some species exhibiting leaf dimorphism (*M. nana, M. reptans, M. tecta,* and *M. tobagensis*) while others are monomorphic (*M. latevagans, M. piloselloides, M. dictyophylla,* and *M. percussa*) (Almeida, 2014; Almeida et al., 2021) (Fig. 1).

The phylogenetic inference by Almeida et al. (2021) revealed potential issues in the circumscription of species within the Scaly clade. For instance, *Microgramma tobagensis*, which has a disjunct distribution between the Andes+northern South America, and the easternmost part of Brazil, was recovered as polyphyletic. A sample from Peru clustered with *M. piloselloides*, while a specimen from eastern Brazil grouped with *M. reptans* (Almeida et al., 2021). Another example is *M. nana*, historically regarded as a variety of *M. tecta*. Despite their close morphological similarity and placement within the same subclade, they are not sister

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

lineages (Almeida et al., 2021). These findings highlight some of the challenges in defining species boundaries within the Scaly clade.

Using a spectral approach, this study aims to understand species boundaries by addressing the following questions: Is FT-NIR spectroscopy an effective tool for testing hypotheses in complex species of ferns, using the Scaly clade of the genus *Microgramma* as a case study? Are there differences in the spectral structure between fertile and sterile fronds, and if so, which is more informative for taxonomic purposes? To this end, we tested the following hypotheses: (I) the current circumscription of some species in the Scaly clade lacks clarity to define lineages because some species are poorly circumscribed; (II) there are significant differences between the spectra of fertile and sterile fronds in dimorphic ferns and less variation in monomorphic ferns.

2 Material and methods

2.1 Sampling

The study was based on vouchers deposited in the following collections: INPA, BHCB, MO, and NY (acronyms according to Thiers, 2024 continuously updated). We analyzed 94 samples from eight species within the Scaly clade (Fig. 1, Table S1), representing geographical and morphological variation within the clade. We followed a thorough protocol to select fronds for the spectral readings, avoiding fronds with damage, brittle, or the presence of fungi or other epiphyllous organisms (Oliveira et al., in prep). All the specimens analyzed in this study were identified and confirmed by the senior author. To test the difference in accuracy between fertile and sterile fronds in monomorphic and dimorphic species, we captured the spectra of three fertile fronds and three sterile fronds from each specimen, whenever available.

2.2 Spectral signature

We collected 1,882 spectral readings in absorbance values, two from each frond surface (abaxial and adaxial), using a Perkin Elmer Spectrum, Frontier MIR+SP10 STD infrared

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

spectrometer (FT-IR). Each measurement lasted 30 seconds, with 32 accumulation measurements to ensure the accuracy and quality of all spectra, at a resolution of 16 cm-1, totaling 2,001 absorbance values per measurement, in the spectral range from 4.000 cm^{-1} to 8.000 cm^{-1} (wavenumber), equivalent to 1.000 to 2.000 nm (wavelength).

The equipment has an automatic performance verification (APV) system, which reduces the frequency of systematic calibrations by continually monitoring performance and alerting to the need for adjustments. Before reading each sample, we performed a background calibration using a white reference sphere made of Spectralon, a fluoropolymer with high reflectance and low absorption. To prevent light scattering, a metal body was placed over the fronds during the spectra capture. All raw data was collected in an environment with a constant and controlled temperature of 18 °C.

2.3 Data pre-processing

After collecting, we ran initial statistical tests to ensure the accuracy and consistency of the spectral data. Any discrepant values (outliers) were manually identified and removed from the dataset. Additionally, we applied standardization using the Standard Normal Variate (SNV) method. This method involves adjusting each spectrum by subtracting the mean and dividing by the standard deviation of each data point (Barnes et al., 1989). This process increases the comparability of the spectra by eliminating dispersion effects and variations in absolute intensity (Barnes et al., 1989). Normalization of spectra using the SNV method was performed using the base R functions 'mean' and 'sd' (Mailund, 2019; R Core Team, 2024).

2.4 Data analysis

We created three models for the spectral tests: Model (I) evaluated spectra from fertile fronds only; Model (II) investigated spectra from sterile fronds only; and Model (III) used a combination of data from fertile and sterile fronds. The data was prepared and manipulated using the *dplyr* package (Mailund, 2019). To explore the variation in raw spectral data in

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

multidimensional space, we performed Principal Component Analysis (PCA) (Kherif & Latypova, 2020) using the *stats* package (R Core Team, 2024).

Partial least squares discriminant analysis (PLS-DA) was performed for each model using the *pls* package (Mevik & Cederkvist, 2004), within the framework provided by the *caret* package (Kuhn, 2008). This analysis involves categorizing samples based on predefined categories, focusing on identifying components that effectively account for variations among the variables in different classes while disregarding noise and uncorrelated variations (Mevik & Cederkvist, 2004). Despite the widespread use of linear discriminant analysis (LDA) in spectral studies (Durgante et al., 2013; Lang et al., 2017; Silva & Hopkins, 2024), its assumptions of normality and homogeneity of covariances (Hastie et al., 2009) are unlikely to apply to spectral data and biological samples.

Two cross-validation techniques were used to assess the performance of the models and their ability to discriminate between species. In the first, K-fold cross-validation, the dataset is divided into K sections or “folds”. In each iteration, the model is trained using K-1 of these folds and validated on the remaining fold. By repeating this process K times, with each fold serving as the validation set exactly once, and averaging the performance results across all iterations, a reliable assessment of the model's performance is obtained (Burman, 1989; Yadav & Shukla, 2016). The second validation was the Leave-One-Out Cross-Validation (LOOCV), which involves selecting a single sample for testing while using the remaining (N-1) samples to train the model. This procedure is repeated N times, so each sample is used once as a test set and (N-1) times as part of the training set (Kohavi, 1995). The final performance of the model is determined by averaging the results of the N iterations (Kohavi, 1995). All analyses were conducted in R version 4.2.3 (R Core Team, 2024).

3 Results

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

The average spectral signature of each species is distinct, with few overlaps (Fig. 2). The PCA including all species (Model III) captured 71.3% of the spectral variation in the first two axes (PCA1=47.7%, PCA2=23.7%), with considerable overlap between species (Fig. 3A). The same overlap was observed when exploring dimorphic and monomorphic species separately (Fig. 3B). Despite overlapping, spectral variation appears to be higher in dimorphic than monomorphic species (3B). This divergence is better seen in the PCA of each species (Fig. 3C), where the dimorphic species (*M. nana*, *M. tecta*, *M. reptans*, and *M. tobagensis*) showed higher variation than the monomorphic species (*M. percussa*, *M. latevagans*, *M. dictyophylla*, and *M. piloselloides*).

In the PCA of model I, using data from fertile fronds only (Fig. 4A), 70.7% of the spectral variation was captured by the first two axes (PCA1=47.3%, PCA2=22.4%). In model II, using sterile fronds (Fig. 4B), the first two axes captured 77.3% of the variation (PCA1=60.8%, PCA2=16.5%). In both models I and II, the spectral overlap between species was quite evident (Fig. 4A-B). The occupation of spectral space by fertile (Model I) and sterile (Model II) leaves varied among species. For *M. dictyophylla*, *M. nana*, and *M. percussa*, fertile leaves (Model I) occupied a smaller spectral space than sterile leaves (Model II). In contrast, *M. reptans* showed the opposite pattern, with fertile leaves (Model I) occupying a broader spectral space than sterile leaves (Model II). The spectral space occupation in *M. latevagans*, *M. piloselloides*, and *M. tecta* did not differ between fertile and sterile leaves (Figs. 4A, 4B).

Based on a PCA of fertile and sterile fronds of all species (Fig. 5A), we found that fertile fronds vary significantly more in the occupation of spectral space than sterile fronds. When examining the variation for each species separately (Fig. 5B), dimorphic species (*M. nana*, *M. reptans*, *M. tecta*, and *M. tobagensis*) showed a higher spectral disparity between fertile and sterile fronds compared to monomorphic species (*M. dictyophylla*, *M. latevagans*, *M. percussa*, and *M. piloselloides*) (Fig. 5B).

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

In the discriminant analysis of the three tested models (Model I - fertile fronds only; Model II - sterile fronds only; and Model III - both fertile and sterile fronds), monomorphic species outperformed dimorphic ones (Table 1). Among the monomorphic species, *M. percussa* stood out with 100% prediction accuracy in all models for both validations (Fig. 6; Table 2). *Microgramma latevagans* and *M. dictyophylla* also performed well, each achieving a prediction accuracy higher than 90%. Among monomorphic species, *M. piloselloides* had the lowest accuracy rates, ranging from 82% to 88% (Fig. 6; Table 2). Despite this, it still outperformed many of the dimorphic species.

Among the dimorphic species, *M. reptans* performed best, with accuracy values ranging from 83% to 95%. *Microgramma tecta* also displayed good results, ranging from 79% to 86% (Fig. 6; Table 2). *Microgramma nana* showed a wide variation, between 74% and 91%, while *M. tobagensis* ranged between 72% and 90% accuracy (Fig. 6; Table 2). Although the dimorphic species showed varied performances, none reached the optimum accuracy achieved by *M. percussa*, which received 100% of the prediction.

The application of pre-processing, specifically removing outliers combined with Standard Normal Variate (SNV) correction (Table 1), demonstrated improved accuracy compared to datasets that only had outliers removed, particularly in Models II and III. Model I, which included fertile fronds only, showed a slight reduction in accuracy after the pre-processing, from 82% to 81% using K-fold cross-validation and from 81% to 77% using leave-one-out cross-validation (LOOCV). Despite this minor reduction, pre-processed data still outperformed raw data (Table 1). For the sterile fronds model (Model II), pre-processing the data resulted in a slight positive change in accuracy. In the K-fold cross-validation, accuracy increased from 73% to 74%, representing a difference of 1.4%, and LOOCV showed an increase from 75% to 76%, with a 1.3% improvement. The combined model (Model III) also improved considerably,

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

with K-fold accuracy increasing from 77% to 78%, a gain of 1.3%, while in the LOOCV, accuracy increased from 77% to 82%, representing a 6.5% improvement.

4 Discussion

In the pioneering study by Paiva et al. (2021), the FT-NIR method was shown to have high accuracy in identifying fern species. However, this is the first study to assess the tool's efficiency on a clade of ferns with recognized species complexes (Almeida et al., 2021). Our results support the usefulness of FT-NIR even in groups with difficult taxonomic delimitation. The accuracy of species discrimination depends on the construction of robust spectral models (Vieira et al., 2021), which, in turn, reflect well-defined circumscriptions, as observed in some species in this study. In addition to species circumscription, variations in the spectral signature of a species can be affected by environmental conditions (e.g., temperature, humidity) (Xu et al., 2019), soil properties (Asner et al., 2014), and developmental stages, such as differences between young and adult fronds due to their chemical composition (Lang et al., 2015; Neuwirthová et al., 2021).

The species tested here are part of the Scaly clade of the genus *Microgramma*, in which morphological overlap and the occurrence of sympatric species make it difficult, in some cases, to precisely define and identify the individuals studied (Almeida, 2014). The average percentage of correct identifications for our three models (sterile fronds, fertile fronds, and both) ranged from 86% to 91% based on the validation methods we used, with an overall average of 88% correct identifications. Only three species (*M. tecta*, *M. nana*, and *M. tobagensis*) had accuracy percentages below 80%.

Microgramma percussa and *M. reptans* were two species that obtained high accuracy results, with 100% and 89% respectively (Table 2). This result was unexpected, given that these species occupy a variety of habitats ranging from open areas to forests, growing on rocks and as epiphytes from lower trunks to the canopy (Almeida, 2014). The fronds of *M. percussa* also

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

present a wide variation in width, length, and shape — a variation that, together with their wide distribution in the Neotropics, from Mexico to southern Brazil, is reflected in its ten heterotypic synonyms (Almeida, 2014). For *Microgramma reptans*, the results are unexpected because this species presents a wide variation in the shape of the sterile fronds, with sterile specimens being hard to distinguish from *M. tobagensis*, for example (Almeida 2014). *Microgramma dictyophylla* and *M. latevagans* also showed high identification accuracy rates, with 93% and 94%, respectively. Although they are morphologically distinct, with variations in leaf shape and size, leaf surface indumentum, and spore type (Almeida, 2014), their distribution areas partially overlap (Almeida, 2014; Lima et al., in prep). *Microgramma latevagans* is endemic to Bolivia and Peru, while *M. dictyophylla* has a wider distribution, covering northern South America. However, they do not occur sympatrically, as *M. latevagans* is found at high elevations (above 2,000 m.a.s.l.), while *M. dictyophylla* grows in lowland forests. Furthermore, the high overall accuracy of their spectral models may reflect good circumscriptions, favored by the specific and striking morphological characteristics of each of them. The few errors in the discriminant analysis related to these two models did not occur between these two species (Fig. 6).

Two morphologically similar species, *M. tobagensis* and *M. piloselloides*, obtained 83% and 84% correct predictions in the PLS-DA (Table 2). The errors might have different explanations. First, the identifications might not be correct, as there is considerable overlap in some of the characteristics used to identify them (e.g., indument and shape of fertile and sterile fronds). Although they mostly do not occur sympatrically (Smith et al., 2018; Almeida et al., 2021), there are populations of *M. piloselloides*, *M. tobagensis*, and *M. reptans* co-occurring in Guatemala and Costa Rica (Almeida, 2014; Lima et al., in prep). Our results showed one incorrect prediction of *M. reptans* as *M. tobagensis* and five incorrect predictions of *M. tobagensis* as *M. reptans* (Fig. 6) — some sterile specimens of *M. tobagensis* make it difficult

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

to distinguish from *M. reptans*. On the other hand, phylogenetic data points to the polyphyly of *M. tobagensis* (Almeida et al., 2021), a conflicting circumscription that may be reflected in the observed spectral errors (Fig. 6). In our investigations, two specimens initially identified as *M.* cf. *tobagensis* were predicted as *M. piloselloides*. After reanalyzing these specimens, the identification was updated and confirmed as *M. piloselloides*, significantly improving the accuracy of the models. We believe that processes such as hybridization and introgression might be playing a role in the evolution of these populations. Adding to the phylogenetic evidence presented by Almeida et al. (2021), spectral data suggest that further studies should focus on a broad sampling of these species populations to test their circumscription.

Microgramma tecta and *M. nana,* which are morphologically similar (Almeida et al., 2021), also had lower predictive identification accuracy, *M. tecta* with an average accuracy of 83% and *M. nana* with 81%, considering all models and validation (Table 2). We believe that the small size of the fertile and sterile fronds, along with the high coverage of trichomes and scales, may have influenced the spectral accuracy. However, no tests have yet been conducted to determine the influence of indument on spectral readings. A fundamental aspect to be considered is the geographic overlap of *M. nana* with other species, such as *M. reptans*, *M. tobagensis*, and *M. dictyophylla*, which raises the possibility of hybridization and introgression between these lineages. Natural hybridization is well documented in vascular plants, especially between phylogenetically close species that co-occur (Liao et al., 2015; Wu et al., 2023). In these sympatry zones, where two or more species coexist under similar environmental conditions, the probability of interspecies crosses increases, generating viable hybrids with a combination of phenotypic traits from the parental species (Sawangproh et al., 2020). Furthermore, introgression – when hybrids backcross with their parental species – can promote the exchange of adaptive genes between populations, increasing genetic variability and potentially favoring local adaptation (Suarez-Gonzalez et al., 2018; Li et al., 2021). These

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

processes, by introducing new genetic traits, can impact the morphological and chemical responses of plants (López-Caamal & Tovar-Sánchez, 2014), creating additional variability between populations. This genetic and phenotypic complexity can theoretically influence the effectiveness of spectral predictive models, although further studies are needed to determine exactly how hybridization and introgression affect the results.

This study is the first to investigate the spectral behavior of functionally different photosynthetic organs, such as the dimorphic fronds of ferns. Heteroblasty and heterophylly are widespread among land plants and have ecological, evolutionary, and taxonomic importance (Zotz et al., 2011). The spectral data suggest that dimorphic species, which have fertile fronds morphologically differing in size and shape from sterile, photosynthetically-only ones, present significant spectral intraindividual variation. Our results also suggest that they have lower prediction percentages when compared to monomorphic species (Fig. 5, Table 1). This relates to the functional and morphological duality of the fronds of dimorphic species (Wagner & Wagner, 1977; Vasco et al., 2013; Watkins et al., 2016), which may introduce higher intrinsic variability into the spectral signature of these species. In contrast, monomorphic species have a single leaf morphotype that performs all functions (Vasco et al., 2013), resulting in less spectral variability within an individual and, therefore, better predictive accuracy.

In fertile fronds, the existence of sori on the abaxial side (Wagner & Wagner, 1977) may have contributed to a good overall accuracy of Model I, although Model III stood out with better accuracy values when considering the species together, without distinguishing between monomorphic or dimorphic ones (Table 1). However, when considering monomorphic and dimorphic species separately, our results indicate better performance for Models II and III (Table 1). Most FT-NIR studies suggest that models such as Model III that combine different data sets, in this case, fertile and sterile fronds, are the best predictive models due to their comprehensive assessment of spectral signatures. For example, Paiva et al. (2021) obtained

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

344 better results with the combination of abaxial and adaxial data. Similarly, Lang et al. (2015)
345 obtained better prediction results by combining data from fronds of young and adult trees. Our
346 results indicate a similar perspective concerning fertile and sterile fronds. The combination of
347 data proved to be better; however, even if data combination can be beneficial, it is not
348 necessarily essential to achieve high accuracy, as observed in our results (Table 1). It is
349 important to consider that the robustness of the models may vary according to the specific
350 context of each study and the characteristics of the species analyzed. Despite testing separate
351 models for fertile and sterile fronds, the morphological variability between these fronds
352 (Watkins et al., 2016) results in different spectral responses in dimorphic species compared to
353 monomorphic ones. This variation occurs because fertile and sterile fronds are subject to
354 additional variations arising from factors such as stage of development, light exposure, and
355 environmental conditions (Moran, 1987; Dalgallo Rocha et al., 2013). This performance
356 underscores the effectiveness and high quality of monomorphic species in the context of our
357 analysis, as evidenced by their higher performance. However, the lower accuracy of dimorphic
358 species does not necessarily diminish its relevance or usability in studies using spectral data. It
359 reflects the complexity and variability associated with the morphofunctional frond duality of
360 these species. We believe that when including dimorphic and monomorphic species in the same
361 set of discriminant and validation data, controlling these variations separately can significantly
362 improve the overall accuracy results of the models as observed in our results (Table 1). Given
363 the diversity of heteroblasty found among ferns (Vasco et al., 2013), an expanded sampling to
364 encompass other lineages is necessary to test whether the pattern we found here is relevant to
365 other groups.
366 This study offers valuable insights into the use of FT-NIR in evidence-based testing of
367 hypotheses in plant species, particularly ferns, which have fewer macromorphological
368 characteristics compared to flowering plants. New methodological multidimensional

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

approaches must be sought, in addition to morphological data, especially when dealing with species complexes. The average overall accuracy of 88% in our validation models was elevated, reflecting good results considering the complexity of the clade. The prediction values are much more significant when applied individually, such as the 100% accuracy for *M. percussa*. Our results reflect in detail the complex relationships of phylogenetically close species, highlighting the usefulness of spectral data for species identification. Finally, we reinforce the importance of a well-planned experimental design for future studies using FT-NIR for other ferns or plant groups. Evaluating and designing effective strategies for utilizing this method will be crucial in employing new approaches and maximizing the potential of spectral data for species identification and analysis.

Author Contributions

NAM: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Visualization; Writing – original draft; Writing – review & editing. MHVO: Data curation; Formal analysis; Writing – original draft; Writing – review & editing. TEA: Conceptualization; Methodology; Project administration; Resources; Validation; Visualization; Writing – original draft; Writing – review & editing

Acknowledgments

This study was partly financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Financing code 001. TEA would like to thank CNPq for grant #317091/2021-2. The authors thank the Federal University of Pernambuco, Herbarium UFP - Geraldo Mariz, and Dr. K. C. Pôrto for the support and logistics with the loans of exsiccates from Brazil and abroad. To the technician, João Carlos da C. Rangel of the Fuels Laboratory linked to the Institute of Research in Oil and Energy for the assistance and provision of the use of the spectrometer equipment; to Dr. E. B. Sessa director of the NY herbarium for releasing the material for spectral analysis; to Dr. M. Sundue, for the authorization to use the image of

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

M. latevagans; and to Dr. F. Durgante, for the knowledge shared and teachings on Spectroscopy.

Conflict of interest

The authors declare no conflict of interest.

References

Alix K, Gérard PR, Schwarzacher T, Heslop-Harris JS. 2017. Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Annals of Botany* 120: 183–194.

Almeida TE, Salino A, Dubuisson JY, Hennequin S. 2021. Insights into long-distance dispersal and ecological and morphological evolution in the fern genus *Microgramma* from phylogenetic inference. *Botanical Journal of the Linnean Society* 196: 294–312.

Almeida TE. 2014. Estudos sistemáticos no gênero *Microgramma* C. Presl (Polypodiaceae - Polypodiopsida). Ph.D. Belo Horizonte: Universidade Federal de Minas Gerais.

Asner GP, Martin RE, Carranza-Jiménez L, Sinca F, Tupayachi R, Anderson CB, Martinez P. 2014. Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region. *New Phytologist* 204: 127–139.

Barnes RJ, Dhano MS, Lister SJ. 1989. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy* 43: 772–777.

Bloesch Z, Nauheimer L, Almeida TE, Crayn D, Field AR. 2022. HybPhaser identifies hybrid evolution in Australian Thelypteridaceae. *Molecular Phylogenetics and Evolution* 173: 107526.

Borges RM. 2008. Plasticity comparisons between plants and animals. *Plant Signaling & Behavior* 3: 367–375.

Burman P. 1989. A Comparative Study of Ordinary Cross-Validation, v-Fold Cross-Validation and the Repeated Learning-Testing Methods. *Biometrika* 76: 503–514.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Cheng S, Zeng W, Wang J, Liu L, Liang H, Kou Y, Wang H, Fan D, Zhang Z. 2021. Species Delimitation of *Asteropyrum* (Ranunculaceae) Based on Morphological, Molecular, and Ecological Variation. *Frontiers in Plant Science* 12: e681864.

Dalgallo Rocha L, Droste A, Gehlen G, Schmitt JL. 2013. Leaf dimorphism of *Microgramma squamulosa* (Polypodiaceae): a qualitative and quantitative analysis focusing on adaptations to epiphytism. *Revista de Biología Tropical* 61: 291–299.

Dayrat B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85: 407–417.

De Queiroz K. 2007. Species Concepts and Species Delimitation. *Systematic Biology* 56: 879–886.

Durgante FM, Higuchi N, Almeida AMR, 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

Edwards DL, Knowles LL. 2014. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B: Biological Sciences* 281: 20132765

Ekrt L, Košnar J, Rothfels CJ, Hanušová K, Hornychn O, Urfus T. 2022. Cytogenetic, geographical, spore type and plastid haplotype data reveal cryptic patterns of species diversity in the cosmopolitan *Cystopteris fragilis* complex (Polypodiopsida: Cystopteridaceae). *Botanical Journal of the Linnean Society* 199: 728–739.

Fernandes AM, Fortini EA, Müller LAC, Batista DS, Vieira LM, Silva PO, Amaral CH, Poethig RS, Otoni WC. 2020. Leaf development stages and ontogenetic changes in passionfruit (*Passiflora edulis* Sims.) are detected by narrowband spectral signal. Journal of *Photochemistry and Photobiology B: Biology* 209: 111931.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

443 Fujiwara T, Liu H, Meza-Torres EI, Morero RE, Vega AJ, Liang Z, Ebihara A, Leitch IJ,
444 Schneider H. 2023. Evolution of genome space occupation in ferns: linking genome diversity
445 and species richness. *Annals of Botany* 131: 59–70
446 Gaem PH, Santos LLD, Andrade A, Vicentini A, Mazine FF. 2020. Three new Amazonian
447 species of *Myrcia* sect. *Mycia* (Myrtaceae) based on morphology and near-infrared
448 spectroscopy. *Phytotaxa* 451: 267–282.
449 Hastie T, Tibshirani R, Friedman J. 2009. Linear Methods for Classification. In: Hastie T,
450 Tibshirani R, Friedman J eds. *The Elements of Statistical Learning: Data Mining, Inference,*
451 *and Prediction*. New York: Springer. 101–137.
452 Haufler CH, Pryer KM, Schuettpelz E, Sessa EB, Farrar DR, Moran R, Schneller JJ, Watkins
453 JE, Windham MD. 2016. Sex and the Single Gametophyte: Revising the Homosporous
454 Vascular Plant Life Cycle in Light of Contemporary Research. *BioScience* 66: 928–937.
455 Heslop-Harrison JS, Schwarzacher T, Liu Q. 2023. Polyploidy: its consequences and enabling
456 role in plant diversification and evolution. *Annals of Botany* 131: 1–10.
457 Hörandl E. 2022. Novel Approaches for Species Concepts and Delimitation in Polyploids and
458 Hybrids. *Plants* 11: 204.
459 Huey RB, Carlson M, Crozier L, Frazier M, Hamilton H, Harley C, Hoang A, Kingsolver, JG.
460 2002. Plants Versus Animals: Do They Deal with Stress in Different Ways? *Integrative and*
461 *Comparative Biology* 42: 415–423.
462 Kherif F, Latypova A. 2020. Principal component analysis. In: Mechelli A, Vieira S eds.
463 *Machine Learning*. Cambridge: Academic Press. 209–225.
464 Kinosian SP, Pearse WD, Wolf PG. 2020. Cryptic diversity in the model fern genus
465 *Ceratopteris* (Pteridaceae). *Molecular Phylogenetics and Evolution* 152: 106938.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Kohavi R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. San Francisco: Morgan Kaufmann Publishers Inc. (IJCAI'95). 1137–1143.

Kuhn M. 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28: 1–26.

Lang C, Almeida DRA, Costa FRC. 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 FRC, Camargo JLC, Durgante FM, Vicentini A. 2015. Near Infrared Spectroscopy Facilitates Rapid Identification of Both Young and Mature Amazonian Tree Species. *PLOS ONE* 10: e0134521.

Li X, Wei G, El-Kassaby YA, Fang Y. 2021. Hybridization and introgression in sympatric and allopatric populations of four oak species. *BMC Plant Biology* 21: 266.

Liao R-L, Ma Y-P, Gong W-C, Chen G, Sun W-B, Zhou R-C, Marczewski T. 2015. Natural hybridization and asymmetric introgression at the distribution margin of two *Buddleja* species with a large overlap. *BMC Plant Biology* 15: 146.

López-Caamal A, Tovar-Sánchez E. 2014. Genetic, morphological, and chemical patterns of plant hybridization. *Revista Chilena de Historia Natural* 87: 16.

Luo J-J, Shang H, Xue Z-Q, Wang Y, Dai X-L, Shen H, Yan Y-H. 2024. Genome-wide data reveal bi-direction and asymmetrical hybridization origin of a fern species *Microlepia matthewii*. *Frontiers in Plant Science* 15: 1392990.

Mailund T. 2019. Manipulating Data Frames: dplyr. In: Mailund T ed. *R Data Science Quick Reference: A Pocket Guide to APIs, Libraries, and Packages*. Berkeley: Apress. 109–160.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Mendez-Reneau JI, Richards JL, Hobbie J, Bollich E, Kooyers NJ, Sigel EM. 2024. Lineage diversification and rampant hybridization among subspecies explain taxonomic confusion in the endemic Hawaiian fern *Polypodium pellucidum*. *American Journal of Botany* e16379.

Mevik BH, Cederkvist HR. 2004. Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR). *Journal of Chemometrics* 18: 422–429.

Moran RC. 1987. Sterile-Fertile Leaf Dimorphy and Evolution of Soral Types in *Polybotrya* (Dryopteridaceae). *Systematic Botany* 12: 617–628.

Neuwirthová E, Kuusk A, Lhotáková Z, Kuusk J, Albrechtová J, Hallik L. 2021. Leaf Age Matters in Remote Sensing: Taking Ground Truth for Spectroscopic Studies in Hemiboreal Deciduous Trees with Continuous Leaf Formation. *Remote Sensing* 13: 1353.

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. 2022. An open and continuously updated fern tree of life. *Frontiers in Plant Science* 13: 909768.

Paiva DNA, Perdiz R de O, Almeida TE. 2021. Using near-infrared spectroscopy to discriminate closely related species: a case study of neotropical ferns. *Journal of Plant Research* 134: 509–520.

Palacio-López K, Beckage B, Scheiner S, Molofsky J. 2015. The ubiquity of phenotypic plasticity in plants: a synthesis. *Ecology and evolution* 5: 3389–3400.

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.

Pinheiro F, Dantas-Queiroz MV, Palma-Silva C. 2018. Plant Species Complexes as Models to Understand Speciation and Evolution: A Review of South American Studies. *Critical Reviews in Plant Sciences* 37: 54–80.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

514 Prananto JA, Minasny B, Weaver T. 2020. Near infrared (NIR) spectroscopy as a rapid and
515 cost-effective method for nutrient analysis of plant leaf tissues. In: Sparks DL ed. *Advances in*
516 *Agronomy*. Amsterdam: Elsevier. 1–49.
517 Prata EMB, Sass C, Rodrigues DP, Domingos FMCB, Specht CD, Damasco G, Ribas CC,
518 Fine PVA, Vicentini A. 2018. Towards integrative taxonomy in Neotropical botany:
519 disentangling the *Pagamea guianensis* species complex (Rubiaceae). *Botanical Journal of the*
520 *Linnean Society* 188: 213–231.
521 Quattrini AM, Wu T, Soong K, Jeng MS, Benayahu Y, McFadden CS. 2019. A next
522 generation approach to species delimitation reveals the role of hybridization in a cryptic
523 species complex of corals. *BMC Evolutionary Biology* 19: 116.
524 R Core Team. 2024. The R Project for Statistical Computing [online]. Available from
525 https://www.r-project.org/ [accessed 01 may 2024].
526 Rodríguez-Fernández JI, Carvalho CJBD, Pasquini C, Lima KMGD, Moura MO, Arízaga
527 GGC. 2011. Barcoding without DNA? Species identification using near infrared
528 spectroscopy. *Zootaxa* 2933: 46–54.
529 Rouhan G, Gaudeul M. 2021. Plant Taxonomy: A Historical Perspective, Current Challenges,
530 and Perspectives. In: Besse, P ed. *Molecular Plant Taxonomy: Methods and Protocols*. New
531 York: Springer US. 1–38.
532 Sandall EL, Maureaud AA, Guralnick R, McGeoch MA, Sica YV, Rogan MS, Booher DB,
533 Edwards R, Franz N, Ingenloff K, Lucas M, Marsh CJ, McGowan J, Pinkert S, Ranipeta A,
534 Uetz P, Wieczorek J, Jetz W. 2023. A globally integrated structure of taxonomy to support
535 biodiversity science and conservation. *Trends in Ecology & Evolution* 38: 1143–1153.
536 Sawangproh W, Lang AS, Hedenäs L, Cronberg N. 2020. Morphological characters and SNP
537 markers suggest hybridization and introgression in sympatric populations of the

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

pleurocarpous mosses *Homalothecium lutescens* and *H. sericeum*. *Organisms Diversity & Evolution* 20: 619–637.

Sigel EM. 2016. Genetic and genomic aspects of hybridization in ferns. *Journal of Systematics and Evolution* 54: 638-655.

Silva GS, Hopkins MJG. 2024. Use of Near Infrared Spectroscopy (NIRS) as a tool to discriminate species of the genus *Dimorphandra* Schott (Leguminosae: Caesalpinioideae). *Scientia Plena* 20: e031201.

Sites, JW, Marshall JC. 2003. Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution* 18: 462–470.

Smith AR, Kessler M, León B, Almeida TE, Jiménez-Pérez I, Lehnert M. 2018. Prodromus of a fern flora for Bolivia. XL. Polypodiaceae. *Phytotaxa* 354: 1–67.

Struck TH, Feder JL, Bendiksby M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson K-H, Liow LH, Nowak MD, Stedje B, Bachmann L, Dimitrov D. 2018. Finding Evolutionary Processes Hidden in Cryptic Species. *Trends in Ecology & Evolution* 33: 153–163.

Suarez-Gonzalez A, Lexer C, Cronk QCB. 2018. Adaptive introgression: a plant perspective. *Biology Letters* 14: 20170688.

Suissa JS, Sundue MA, Testo WL. 2021. Mountains, climate and niche heterogeneity explain global patterns of fern diversity. *Journal of Biogeography* 48: 1296–1308.

Thiers BM. 2024. Index Herbariorum - The William & Lynda Steere Herbarium [online]. Available from https://sweetgum.nybg.org/science/ih/ [accessed 01 may 2024].

Vasco A, Moran RC, Ambrose BA. 2013. The evolution, morphology, and development of fern leaves. *Frontiers in Plant Science* 4: 345.

Vasconcelos C, Ferraz IDK, Adrianzén MU, Camargo JLC, Terra-Araujo H. 2021. *Chromolucuma brevipedicellata* (Sapotaceae, Chrysophylloideae), a new tree species from central Amazonia, Brazil. *Brittonia* 73: 211–219. Vieira LS, Assis C, de Queiroz, MELR,

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

563 Neves AA, de Oliveira AF. 2021. Building robust models for identification of adulteration in
564 olive oil using FT-NIR, PLS-DA and variable selection. *Food Chemistry* 345: 128866.
565 Vieira LS, Assis C, de Queiroz MELR, Neves AA, de Oliveira AF. 2021. Building robust
566 models for identification of adulteration in olive oil using FT-NIR, PLS-DA and variable
567 selection. *Food Chemistry* 345: 128866.
568 Wagner WH, Wagner FS. 1977. Fertile–sterile leaf dimorphy in ferns. *Gardens Bulletin*
569 *(Singapore)* 30: 251–268.
570 Watkins JE, Churchill AC, Holbrook NM. 2016. A site for sori: Ecophysiology of fertile-
571 sterile leaf dimorphy in ferns. *American Journal of Botany* 103: 845–855.
572 Wei Z, Xia Z, Shu J, Shang H, Maxwell SJ, Chen L, Zhou X, Xi W, Adjie B, Yuan Q, Cao J,
573 Yan Y. 2022. Phylogeny and Taxonomy on Cryptic Species of Forked Ferns of Asia.
574 *Frontiers in Plant Science* 12: 748562.
575 Wiens JJ. 2007. Species Delimitation: New Approaches for Discovering Diversity. *Systematic*
576 *Biology* 6: 875–878.
577 Will KW, Mishler BD, Wheeler QD. 2005. The Perils of DNA Barcoding and the Need for
578 Integrative Taxonomy. *Systematic Biology* 54: 844–851.
579 Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The
580 frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of*
581 *Sciences of the United States of America* 106: 13875–13879.
582 Workman J, Weyer L. 2007. Practical Guide to Interpretive Near-Infrared Spectroscopy. Boca
583 Raton: CRC Press.
584 Wu Y, Yu X, Tang W, Yang W, Fu Q, Zheng Y, Zhang C. 2023. Morphological and
585 molecular evidence for natural hybridization between *Sorbus pohuashanensis* and *S. discolor*
586 (Rosaceae). *Journal of Forestry Research* 35: 25.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

587 Xu H, Yu P, Fu X, Ying Y. 2009. On-site variety discrimination of tomato plant using visible-
588 near infrared reflectance spectroscopy. *Journal of Zhejiang University SCIENCE B* 10: 126–
589 132.
590 Xu X, Xie L,Ying Y. 2019. Factors influencing near infrared spectroscopy analysis of agro-
591 products: a review. *Frontiers of Agricultural Science and Engineering* 6: 105–115.
592 Yadav S, Shukla S. 2016. Analysis of k-Fold Cross-Validation over Hold-Out Validation on
593 Colossal Datasets for Quality Classification. In: 2016 IEEE 6th International Conference on
594 Advanced Computing (IACC), 78–83. Available from
595 https://ieeexplore.ieee.org/document/7544814 [accessed 13 July 2024].
596 Yeates DK, Seago A, Nelson L, Cameron SL, Joseph L, Trueman JWH. 2011. Integrative
597 taxonomy, or iterative taxonomy? *Systematic Entomology* 36: 209–217.
598 Yi H, Dong S, Yang L, Wang J, Kidner C, Kang M. 2023. Genome-wide data reveal cryptic
599 diversity and hybridization in a group of tree ferns. *Molecular Phylogenetics and Evolution*
600 184: 107801.
601 Zahir SADM, Omar AF, Jamlos MF, Azmi MAM, Muncan J. 2022. A review of visible and
602 near-infrared (Vis-NIR) spectroscopy application in plant stress detection. *Sensors and*
603 *Actuators A: Physical* 338: 113468.
604 Zotz G, Wilhelm K, Becker A. 2011. Heteroblasty—A Review. *Botanical Review* 77: 109–
605 151.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Tables

Table 1 Accuracy (%) of models tested in validations and Confidence Interval (CI%) before and after pre-processing.

Data	Raw data		Outlier removal		Outlier removal + Standard Normal Variate (SNV)		Monomorphic species only		Dimorphic species only	
Models /Validation	K-fold	LOOCV	K-fold	LOOCV	K-fold	LOOCV	K-fold	LOOCV	K-fold	LOOCV
Fertile	77% (69-85%)	77% (69-84%)	82% (73-89%)	81% (73-88%)	81% (71-88%)	77% (67-85%)	95% (84-99%)	98% (88-100%)	76% (61-87%)	73% (57-86%)
Sterile	75% (66-84%)	78% (70-86%)	73% (65-82%)	75% (66-83%)	74% (64-83%)	76% (66-84%)	100% (90-100%)	100% (92-100%)	76% (60-88%)	83% (69-93%)
both	77% (69-84%)	76% (68-84%)	77% (69-85%)	77% (68-85%)	78% (68-86)	82% (73-89%)	100% (92-100%)	100% (90-100%)	77% (61-89%)	83% (68-93%)

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Table 2 Number of species and hit rates (balanced accuracy%) for the tested models and functional types of fronds of the Scaly *Microgramma* clade (Polypodiaceae).

Species	Number of species	Model I - fertile		Model II - sterile		Model III - both		Average accuracy values
		K-fold	LOOCV	K-fold	LOOCV	K-fold	LOOCV	
Monomorphic	44	-	-	-	-	-	-	-
M. dictyophylla	08	93%	99%	90%	90%	90%	94%	93%
M. latevagans	07	99%	94%	93%	93%	94%	94%	94%
M. percussa	11	100%	100%	100%	100%	100%	100%	100%
M. piloselloides	19	88%	82%	84%	82%	83%	85%	84%
Dimorphic	50	-	-	-	-	-	-	-
M. reptans	14	87%	86%	94%	95%	83%	87%	89%
M. nana	11	80%	74%	77%	77%	90%	91%	81%
M. tecta	11	83%	84%	79%	86%	81%	86%	83%
M. tobagensis	13	88%	83%	72%	75%	88%	90%	83%

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Figures

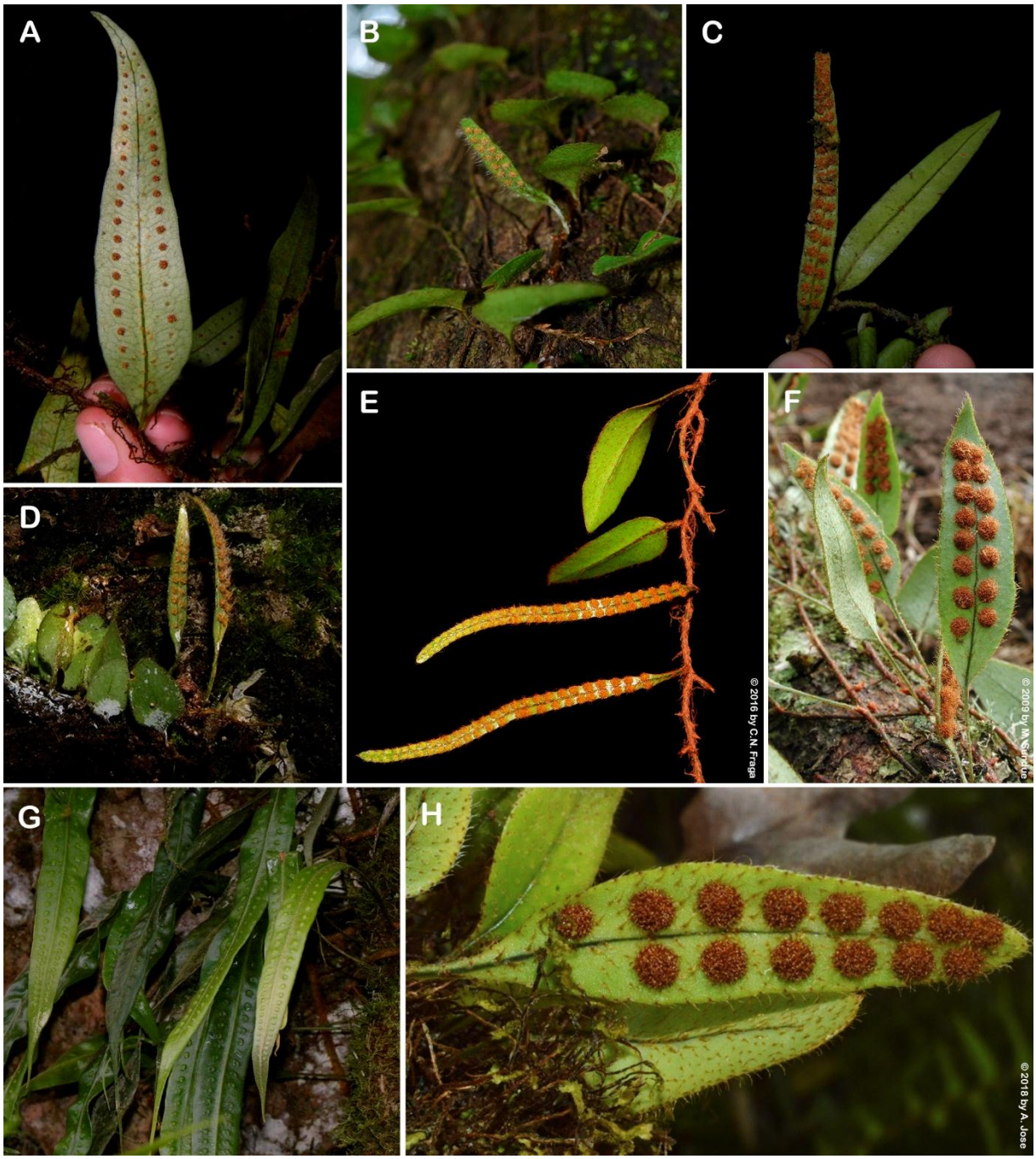

Fig. 1. *Microgramma* species from the Scaly clade (Polypodiaceae). **A,** *M. dictyophylla*. **B,** *M. nana*. **C,** *M. tobagensis*. **D,** *M. tecta*. **E,** *M. reptans* (© by 2016 C.N. Fraga). **F,** *M. latevagans* (© by 2009 M. Sundue). **G,** *M. percussa*. **H,** *M. piloselloides* (© by 2018 A. Jose).

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

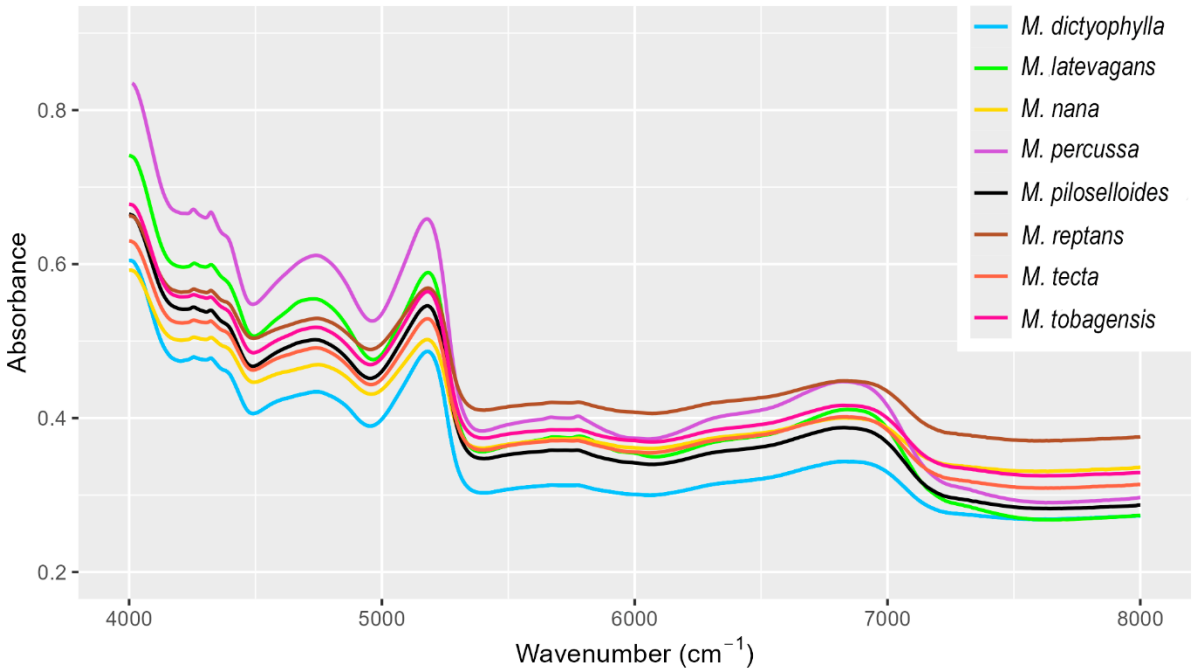

Fig. 2. Spectral signature (average data) for eight *Microgramma* species from the Scaly clade.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Fig. 3. Principal component analysis (PCA). **A,** Complete dataset showing the eight tested species: *Microgramma dictyophylla*, *M. latevagans*, *M. nana*, *M. percussa*, *M. piloselloides*, *M. reptans*, and *M. tobagensis*. **B,** PCA showing specimens categorized by the two functional types of fronds, dimorphic and monomorphic. **C,** Same PCA, showing the tested species separately.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

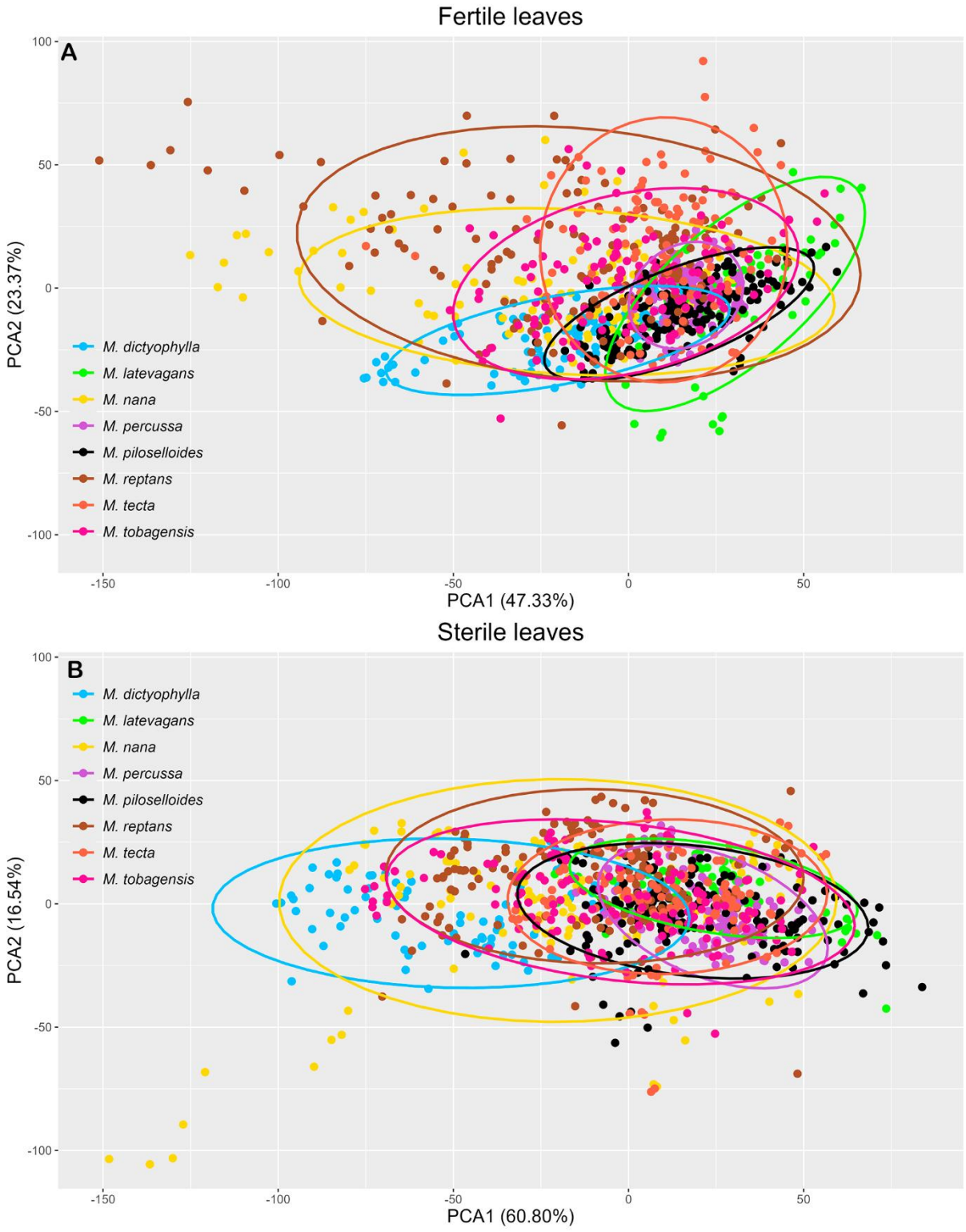

Fig. 4. Principal component analysis (PCA). **A,** Dataset of the eight species plotted using the fertile leaf category. **B,** Dataset of the eight species plotted using the sterile leaf category.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

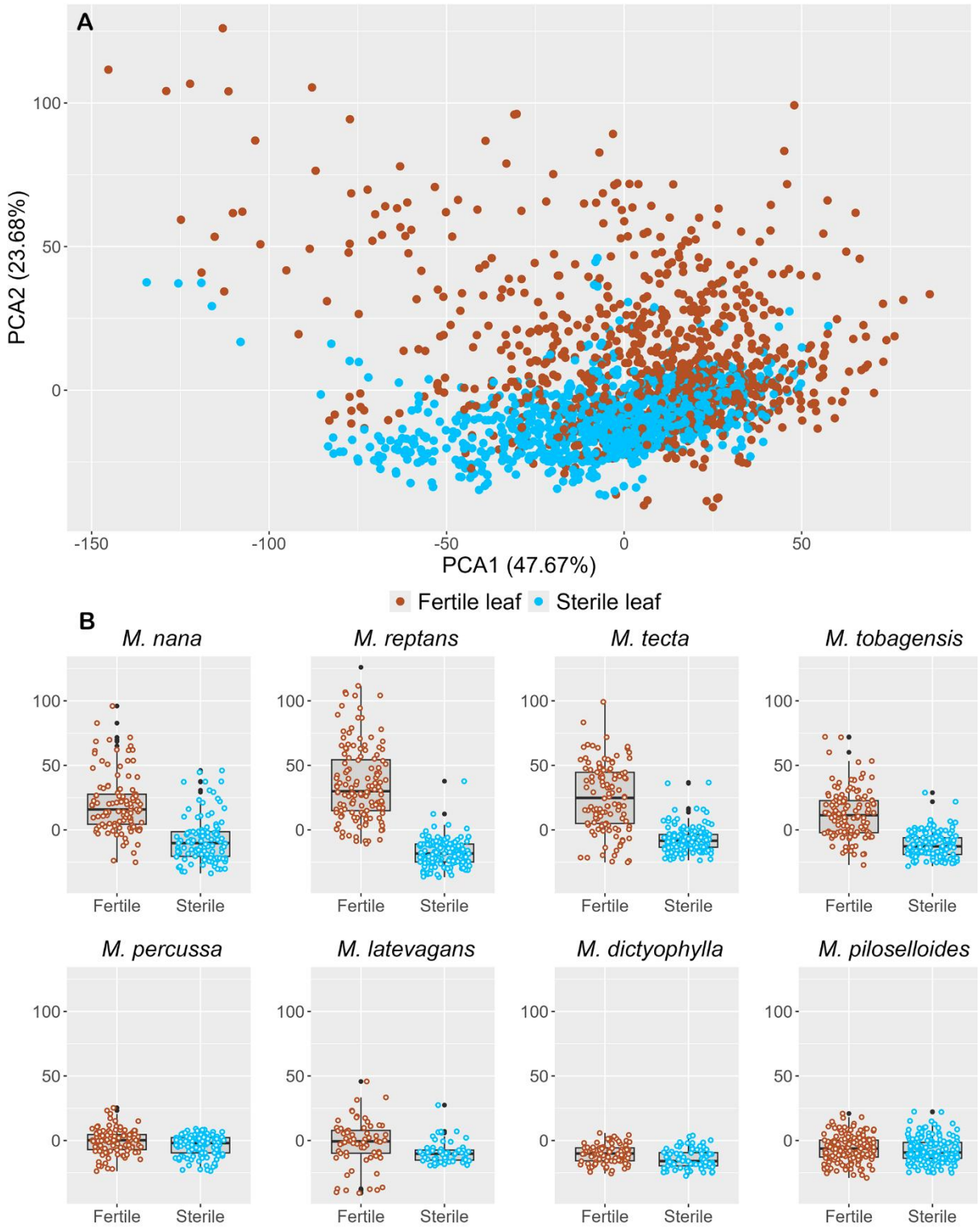

Fig. 5. Principal component analysis (PCA) plot. **A,** Dataset of the eight species plotted, using the fertile and sterile leaf category. **B,** Fertile and sterile leaf category dataset by species.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

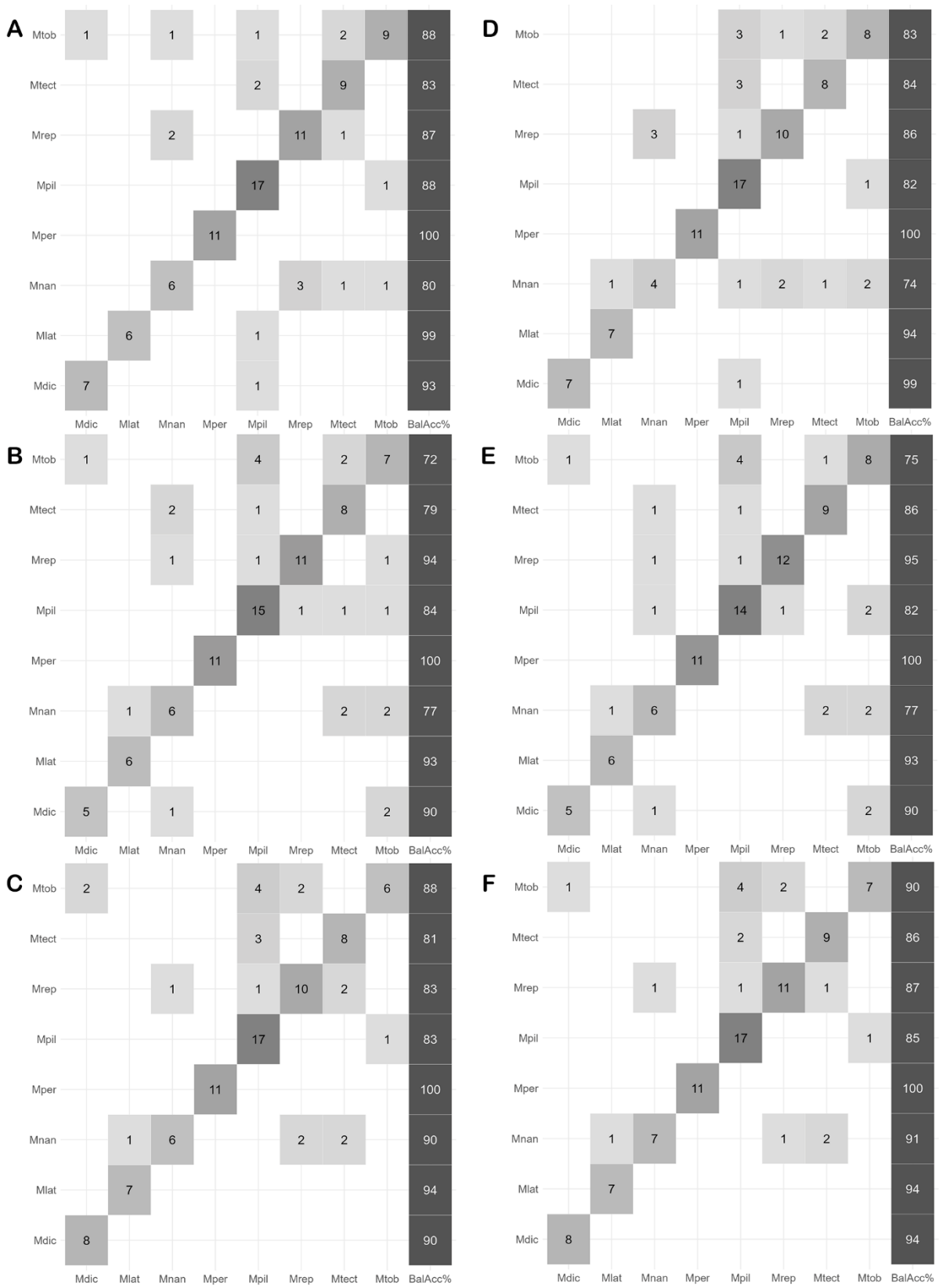

Fig. 6. Confusion matrices resulting from partial least squares discriminant analysis (PLS-DA) for K-fold and LOOCV validations. **A,** K-fold validation for fertile fronds. **B,** K-fold validation for fertile fronds. **C,** K-fold validation for combined fronds. **D,** LOOCV validation for fertile fronds. **E,** LOOCV validation for fertile fronds. **F,** LOOCV validation for combined fronds.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Supplementary electronic materials

Title: Spectral signatures as evidence to test hypotheses in plant species complexes

Authors: Niksoney Azevedo Mendonça, Marise Helen Vale de Oliveira and Thaís Elias Almeida

Journal: Journal of Systematics and Evolution

Corresponding author: Niksoney Azevedo Mendonça (Universidade Federal de Pernambuco, Programa de Pós-graduação em Biologia Vegetal). Tel: +55-98-991213636 E-mail: niksoney.azevedo@ufpe.br

Table S1 – Specimens used for spectral data capture. The acronym Herbaria follows Thiers (2024 continually updated: https://sweetgum.nybg.org/science/ih/).

Table S1

Species	**Country**	**Voucher**
Microgramma dictyophylla	Brazil	Vida 739 (BHCB145204)
Microgramma dictyophylla	Brazil	Almeida 2230 (BHCB136581)
Microgramma dictyophylla	Brazil	Almeida 2605 (BHCB144728)
Microgramma dictyophylla	Brazil	Cowan 38507 (NY00880074)
Microgramma dictyophylla	Ecuador	Gilmartin 320 (MO1799919)
Microgramma dictyophylla	Guyana	Maguire 40547 (NY03961499)

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Microgramma dictyophylla	Peru	Foster 7419 (MO3232987)
Microgramma dictyophylla	Venezuela	Liesner 24298 (NY03961515)
Microgramma latevagans	Bolívia	Smith 13487 (MO4008613)
Microgramma latevagans	Bolívia	Rodriguez 1263 (MO6145673)
Microgramma latevagans	Bolívia	Antezana 1438 (MO04857892)
Microgramma latevagans	Bolívia	Lewis 881089 (MO4008174)
Microgramma latevagans	Peru	Sánchez Vega 5979 (MO3292118)
Microgramma latevagans	Peru	Dillon 6092 (NY03230430)
Microgramma latevagans	Peru	Bennett 2601 (NY03349971)
Microgramma nana	Brazil	Almeida 2732 (BHCB149816)
Microgramma nana	Brazil	Silva 168 (INPA113289)
Microgramma nana	Brazil	Carvalho UAT83 (INPA226341)
Microgramma nana	Brazil	Almeida 2632 (BHCB144755)
Microgramma nana	Costa Rica	Grayum 4884 (MO3324161)
Microgramma nana	Ecuador	Aulestia 128 (MO6016797)
Microgramma nana	Ecuador	Pérez 1185 (NY03962278)
Microgramma nana	México	Campos Villanueva 1011 (MO5069443)
Microgramma nana	Suriname	Maguire 24002 (MO1312664)
Microgramma nana	Suriname	Herrera 10088 (MO6069412)
Microgramma nana	Venezuela	Davidse 3042 (MO2982712)
Microgramma percussa	Brazil	Fraga 3104 (BHCB150254)
Microgramma percussa	Brazil	Dittrich 842 (BHCB64514)
Microgramma percussa	Brazil	Barreto 2595 (BHCB154070)
Microgramma percussa	Brazil	Almeida 372 (BHCB101385)
Microgramma percussa	Brazil	Salino 1988 (BHCB29357)

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Microgramma percussa	Brazil	Madison 618 (INPA85192)
Microgramma percussa	Colombia	Croat 79621 (MO6027011)
Microgramma percussa	Colombia	Luteyn 10469 (NY03961982)
Microgramma percussa	México	Hernández 445 (NY03961228)
Microgramma percussa	Panama	McDaniel 8077 (MO6576856)
Microgramma percussa	Peru	Campos 3991 (MO5302244)
Microgramma piloselloides	Bonaire	Boom 11057 (NY02099903)
Microgramma piloselloides	Costa Rica	Rojas 2067 (MO5900546)
Microgramma piloselloides	Cuba	Underwood 831 (NY01842269)
Microgramma piloselloides	Cuba	Britton 5165 (NY01842276)
Microgramma piloselloides	Dominica	Chambers 2501 (NY01842425)
Microgramma piloselloides	Grenada	Broadway s.n. (MO5446677)
Microgramma piloselloides	Jamaica	s.c/s.n. (MO5462147)
Microgramma piloselloides	Jamaica	s.c/s.n. (MO5462145)
Microgramma piloselloides	Jamaica	Hitchcock s.n. (MO5446665)
Microgramma piloselloides	Jamaica	s.c/s.n. (MO5446680)
Microgramma piloselloides	Puerto Rico	Ahlquist 45 (BHCB173388)
Microgramma piloselloides	Puerto Rico	Britton 2062 (NY00989767)
Microgramma piloselloides	Puerto Rico	Underwood 11 (NY00989773)
Microgramma piloselloides	Puerto Rico	Shafer 3139 (NY00989780)
Microgramma piloselloides	Puerto Rico	Vincent 15355 (NY02150502)
Microgramma piloselloides	República do Haiti	Nash 217 (NY01842393)
Microgramma piloselloides	Republica Dominicana	Mejía 7508 (MO3859988)
Microgramma piloselloides	Republica Dominicana	Zanoni 28119 (NY01842360)

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Microgramma piloselloides	St Vicent and The Grenadines	Badger s.n. (MO5484127)
Microgramma reptans	Brazil	Daly 1352 (INPA118203)
Microgramma reptans	Brazil	Matos 11468 (CEPEC15838)
Microgramma reptans	Brazil	Fraga 3090 (BHCB150244)
Microgramma reptans	Brazil	Fraga 2971 (BHCB143826)
Microgramma reptans	Colombia	Molina 18c716 (MO1626606)
Microgramma reptans	Costa Rica	Lesica 4136 (MO3416567)
Microgramma reptans	Ecuador	Aulestia 1560 (MO05055185)
Microgramma reptans	Guyana	MC Dowell 4809 (MO04621215)
Microgramma reptans	Nicaragua	Urbina 70 (MO6710204)
Microgramma reptans	Panama	Salino 15879 (BHCB173590)
Microgramma reptans	Panama	Stern 646 (MO1817771)
Microgramma reptans	Panama	Kennedy 3232 (MO3014609)
Microgramma reptans	Peru	Schunke-Vigo 3469 (INPA133051)
Microgramma reptans	Peru	King 451 (INPA123964)
Microgramma tecta	Brazil	Krieger 8862 (BHCB165025)
Microgramma tecta	Brazil	Salino s.n. (BHCB43418)
Microgramma tecta	Brazil	Salino 10145 (BHCB91558)
Microgramma tecta	Brazil	Salino 13779 (BHCB124333)
Microgramma tecta	Brazil	Mota 3160 (BHCB105182)
Microgramma tecta	Brazil	Schmitt 334 (BHCB142994)
Microgramma tecta	Brazil	Kollmann 4051 (BHCB108513)
Microgramma tecta	Brazil	Matos 415 (BHCB97283)
Microgramma tecta	Brazil	Salino 13595 (BHCB122753)

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.10972

Microgramma tecta	Brazil	Salino 8637 (BHCB81163)
Microgramma tecta	Brazil	Cadorin 2121 (BHCB142972)
Microgramma tobagensis	Brazil	Thomas 9304 (MO04895073)
Microgramma tobagensis	Brazil	Mori 11468 (NY00674449)
Microgramma tobagensis	Brazil	Amorim 4205 (NY01241079)
Microgramma tobagensis	Brazil	Matos 1577 (NY02064119)
Microgramma tobagensis	Ecuador	Gudiño Jara 2256 (MO04911176)
Microgramma tobagensis	Ecuador	Clark 575 (MO5149584)
Microgramma tobagensis	Ecuador	Vargas lópez 5138 (MO6389474)
Microgramma tobagensis	Guatemala	Türckheim 125 (NY03961400)
Microgramma tobagensis	Peru	Klug 2539 (NY03966108)
Microgramma tobagensis	Trinidad and Tobago	Broadway 9959 (NY02027319)
Microgramma tobagensis	Trinidad and Tobago	Jermy 2870 (NY02027320)
Microgramma tobagensis	Trinidad and Tobago	Jermy 2680 (MO04642633)
Microgramma tobagensis	Venezuela	Steyermark 61888 (NY03966137)

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença Creative Commons CC-BY.
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.