Contents lists available at ScienceDirect

# Forest Ecology and Management

# Tree species delimitation in tropical forest inventories: Perspectives from a taxonomically challenging case study

Paulo Henrique Gaem [a,*], Ana Andrade [b], Fiorella Fernanda Mazine [a], Alberto Vicentini [b,c]

[a] Grupo de Pesquisa em Ecologia, Sistemática e Conservação de Recursos Naturais, Universidade Federal de São Carlos, Campus Sorocaba, Rodovia João Leme dos Santos (SP-264), km 110, 18052-780 Sorocaba, SP, Brazil
[b] Projeto Dinâmica Biológica de Fragmentos Florestais, Instituto Nacional de Pesquisas da Amazônia, Avenida André Araújo, 2936, 69067-375 Manaus, AM, Brazil
[c] Laboratório de Botânica Amazônica, Coordenação Dinâmica Ambiental, Instituto Nacional de Pesquisas da Amazônia, Avenida André Araújo, 2936, 69067-375 Manaus, AM, Brazil

## ARTICLE INFO

## ABSTRACT

Forest conservation and management programs rely on precise knowledge of local tree diversity. Distinguishing tree species is a major concern in tropical forest inventories as morphology alone may not be sufficient for species identifications in those highly diverse environments. Therefore, a multi-evidence approach is desirable to avoid subjective interpretations. We tested the contribution of a low-cost integrative protocol for delimiting sympatric species of the taxonomically challenging genus *Myrcia* (Myrtaceae) in the central Amazon region. We classified specimens into local species by morphotyping them and then validated our hypotheses using two additional sources of evidence: i) near-infrared spectroscopic data, consisting of 1557 absorbance values obtained from dried leaves; and, ii) morphometric data of leaf blade contours, translated into Fourier descriptors and summarized by principal component analysis (PCA), together with four additional leaf characters. The morphotypes were validated using linear discriminant analyses (LDAs) of the raw NIR spectra, stepwise-selected spectral regions, PCA-reduced spectra, as well as the morphometric data. A taxonomic name was assigned to each validated species based on comparisons with herbarium specimens and/or the technical literature. We delimited 38 species of *Myrcia,* and our hypotheses were well-supported by the LDAs (81–99% accuracy), evidencing that inexpensive tools can be effectively used to discriminate species in large-scale projects, and that integrative approaches are fundamental in that regard. Although vegetative traits were sufficient for species discrimination, fertile samples were crucial for obtaining taxonomic names for them. Thirteen species delimited by us belong to four species complexes, each treated under single species names in the current systematics of the genus. We therefore argue that species delimitations prior to individual identifications are essential for robust species definition in forest plot research because local species do not necessarily match global circumscriptions. This type of approach may greatly alter how local community structures are viewed and consequently modify the results of downstream analyses of ecological studies. We also emphasize that local species delimitations may contribute to the field of taxonomy by identifying potential inconsistencies in global species definitions.

## 1. Introduction

A full understanding of tropical forest structure is essential to ecological studies focusing on conservation and management activities, and will greatly depend on accurately distinguishing plant species, especially trees. Species delimitations are important as those taxa may have different roles and functions in ecosystems (Chapin et al., 1996), although such delimitations are never simple tasks in these highly diverse environments. While macro-morphology (hereinafter simply 'morphology') is by far the most direct and least expensive operational criterion used for that purpose, it may be of very limited use in some cases: abrupt morphological differences between individuals may not always be sufficient to accurately infer that they represent different species, as the observed differences may only indicate phenotypic plasticity in response to environmental or ontogenetic influences (Marenco et al., 2017), and may especially complicate species delimitations when juveniles and adults are intermixed. In addition, morphology can often be confusing in cases of cryptic variation,
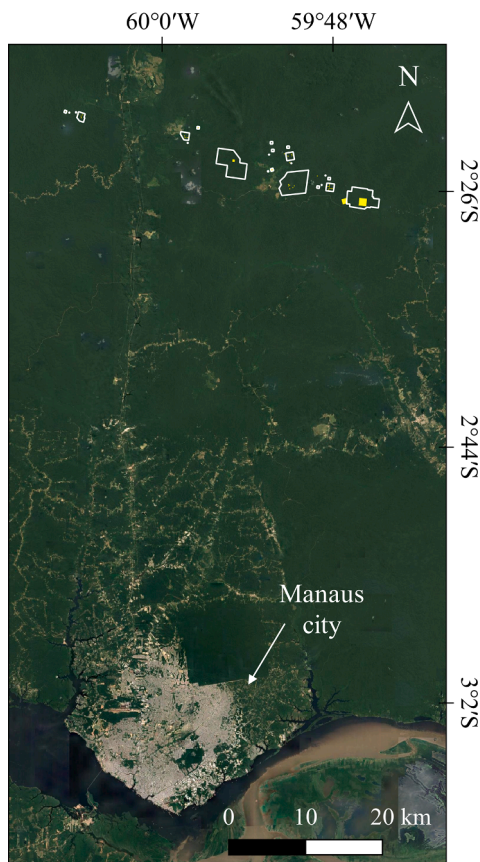
---

**Fig. 1.** Permanent plots of the Biological Dynamics of Forest Fragments Project (yellow polygons) within the BDFFP Area of Relevant Ecological Interest (transparent polygons), in Amazonas state, Brazil. [SIRGAS 2000 GCS] (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hybridization, or convergent evolution (e.g., Borba et al., 2002; Crawford & Stuessy, 2016), potentially leading to uncertainties when treated alone. It is therefore relatively common to encounter undescribed species or to collect two or more coexisting species that fall into a single taxonomic name with wide morphological circumscription (e.g., Esteves & Vicentini, 2013), and even presumably 'good species' can turn out to comprise numerous old-divergent lineages, with non-obvious morphological distinctions (Prata et al., 2018; Damasco et al., 2021).

Local species delimitations in tropical forest inventories should be easier given the sympatric context and the absence of variation due to geography (see the *adminensional species* of Mayr, 1992), but it is actually affected by subjective interpretations of morphology (Gomes et al., 2013). Current plant taxonomy depends heavily on the analysis of reproductive materials, but these are difficult to obtain in such environments; consequently, most samples in inventories are obtained sterile, and the lack of flowers or fruits make it difficult to decide whether some morphotypes are derived from intraspecific variations or different species. Moreover, assigning taxonomic names to local species is difficult due to the poor status of the botanical knowledge in the Amazon (Hopkins, 2007). Hence, defining local species in forest inventories and finding a taxonomic name for them are extremely difficult tasks in Amazonian forest research.

Approaches combining traditional morphology with morphometric evidence as well as modern computational methods offer the promise of refining plant species richness assessments (Draper et al., 2020). Although DNA sequencing represents the best source of information, it may be impractical in large-scale surveys due to its high costs. Low-cost phenotypic characterizations, however, offer the possibility of

distinguishing closely related taxa, and near-infrared spectroscopy (NIRS) applied to plant tissue, for example, is capable of measuring molecular vibrations and bending, yielding complex absorbance patterns that allow the generation of predictive models of internal composition and structure (Durgante et al., 2013). In parallel, geometric morphometrics can transform leaf shapes into mathematical information, making it possible to treat those patterns objectively (Viscosi & Cardini, 2011).

Here, we combined traditional morphological approaches with NIRS and leaf shape geometries to study species of the genus *Myrcia* (Myrtaceae) growing in permanent forest plots of the Biological Dynamics of Forest Fragments Project (BDFFP), in Central Amazon (Fig. 1). Most of these plots were implemented in the early 1980's to study the effects of fragmentation of tropical forests (Laurance et al., 2011). The predominant local vegetation there is upland *terra firme* forest on clayey soils, and this site is among the most diverse forests on Earth (Oliveira & Mori, 1999; Duque et al., 2017).

Approximately 290,000 trees have been systematically tagged and collected in the BDFFP plots, belonging to ca. 2450 species, of which ca. 30% are morphospecies without formal binomial names. Although all these trees have been collected, defining local species (i.e. morphotyping, or grouping morphologically similar specimens) has been a convoluted task at BDFFP due to the high diversity, the coexistence of many closely related and similar species (Oliveira & Mori, 1999), and the inherent subjectivity of the morphotyping process (Gomes et al., 2013). The genus *Myrcia* is one of the most problematic in terms of delimitation of species in the BDFFP flora, as pointed out by Gomes et al. (2013) for the *Myrcia splendens* (Sw.) DC. Amazonian complex (treated in that work as '*M. fallax-deflexa*' and *M. magnoliifolia* DC.), an issue also present throughout the geographical range of this taxon (Lucas et al., 2011). Considering these problems and the richness of the genus in the study area, a careful examination of the boundaries of the species of *Myrcia* occuring in the BDFFP plots was undertaken aiming at the following question: can a low-cost integrative protocol contribute to the task of circumscribing species of a taxonomically challenging group of plants in a forest inventory?

## 2. Material and methods

### 2.1. Sampling

Fertile and sterile specimens belonging to *Myrcia* collected in the study site and deposited in the BDFFP collection were used in this study, including samples of both juveniles/treelets (1–10 cm diameter at breast height, DBH) and trees (DBH > 10 cm). We initially reviewed the collections and reclassified the specimens into local species by morphotyping. Those initial species hypotheses were used to select 375 specimens from a total of 3384 for taking measurements. We also incorporated 12 fertile specimens from the INPA herbarium that had been collected near the study site to complement our sampling in two cases: taxa of doubtful identity composed of sterile samples only [*Myrcia crebra* (McVaugh) A.R.Lourenço & E.Lucas and *M. gigas* McVaugh] and locally rare species with very limited sampling [*M. intonsa* (McVaugh) B. Holst, *M. neospeciosa* A.R.Lourenço & E.Lucas, and *M. pyrifolia* (Desv.) Nied.]. Supplementary File 1 details the samples used per species.

We used leaf-based NIRS and morphometric data to test our species hypotheses. Spectral data was obtained using an Antaris™ II FT-NIR Analyzer (Thermo Fisher Scientific). Dried leaves were placed over a radiation emitter, yielding 1557 absorbance values in the range 4000–10,000 cm$^{-1}$ per reading. A black body was placed over the leaves during the readings to prevent light dissipation. We measured the absorbance spectra of both the adaxial and abaxial surfaces of two well-developed healthy leaves per sample, twice on each surface, and calculated the mean spectral values for each specimen. For leaf shape data we scanned two complete leaves per specimen and posteriorly translated them into Fourier descriptors of leaf contours using Shape

**Table 1**

Morphometric characters measured from leaves of samples for application in discriminant analyses.

| Character | Type | Variation states |
|---|---|---|
| Adaxial disposition of midvein and lateral veins relative to leaf blade | Qualitative | Impressed, flat, raised |
| Indumentum of leaf blade on both surfaces | Qualitative | Absent, glabrate, sparse, dense, total |
| Number of pairs of lateral veins | Quantitative and discrete | – |
| Proportion of leaf blade width occupied by the space between the inner marginal vein and blade margin | Quantitative and continuous (mm) | – |

(Iwata & Ukai, 2002), a software that generates shape data summarized into principal components (PCs), each informing a particular aspect of the variation. We conducted visual inspections of the output shape data to select only taxonomically informative PCs, excluding those that captured noise or symmetry effects between two opposite leaves positioned at the same node. We also calculated leaf area using the *EBImage* package (Pau et al., 2010) of the R environment (R Core Team, 2020) and measured four other leaf characters (Table 1). The data obtained is available as Supplementary File 2.

## 2.2. Analyses

All analyses were performed in the R environment. As spectral data result in huge numbers of highly correlated variables (greater than the number of samples, in our case), we performed two data dimensionality reduction procedures. The first involved a stepwise variable selection, using the *stepclass* function of the *klaR* package (Weihs et al., 2005), that tests the contribution of variables to discriminatory functions by including and excluding them from the model one at a time. We then reduced the raw NIRS data using principal component analyses (PCA), which is advantageous for yielding non-collinear descriptors. We chose numbers of variables that corresponded to 1/3 of the number of samples in both cases, as recommended by Williams and Titus (1988).

We performed linear discriminant analyses (LDAs) with the three versions of the NIRS dataset: the raw spectra, the stepwise-selected regions, and the PCA-reduced spectra. We also tested the discriminatory power of the morphometric data, both alone and combined with each version of the NIRS data, totaling seven tested datasets. We implemented leave-one-out with cross validation (LOOCV) LDAs to test the robustness of our species hypotheses by predicting the identity of each specimen based on the identity of all other specimens; hence, individual observations were used as validation data, and all other observations were used as training data. To check if our tested evidence indeed translated
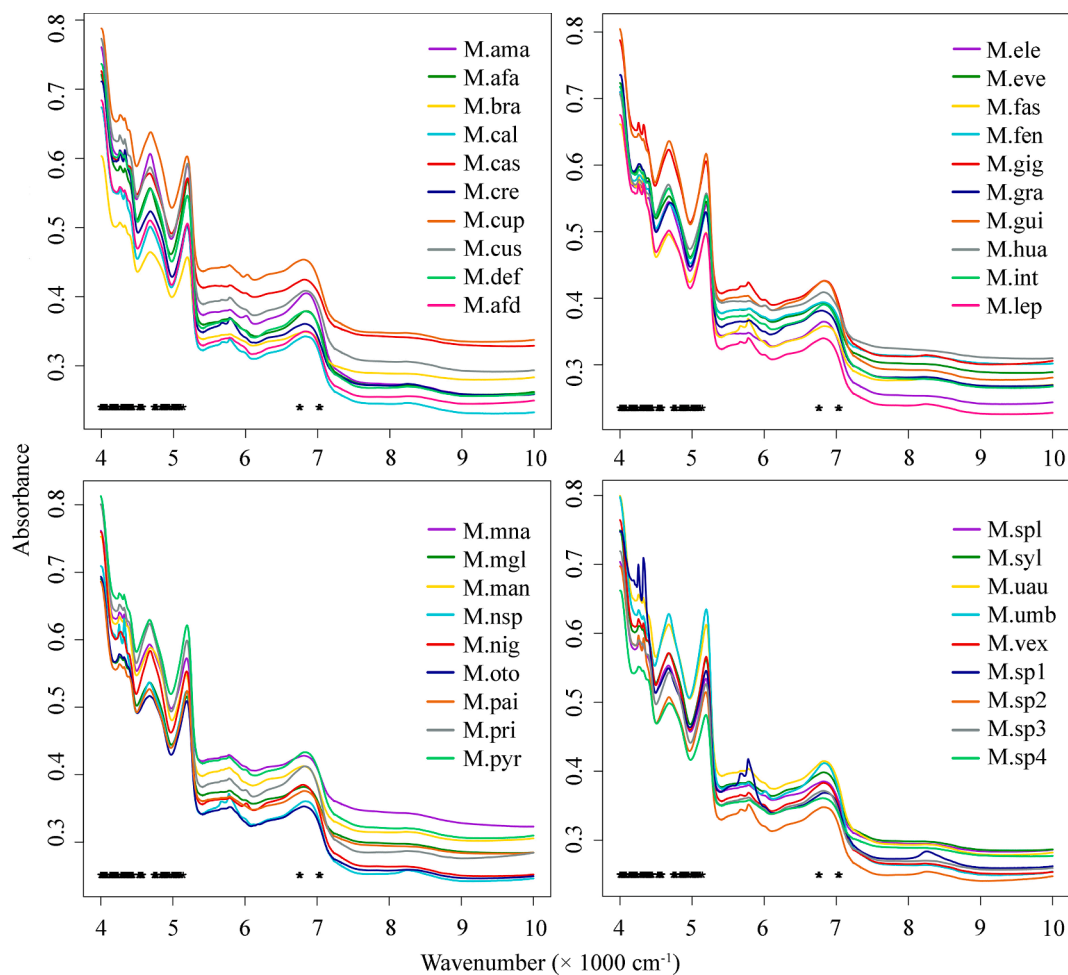


**Fig. 2.** Average near-infrared spectrum of each species. Lines represent spectra composed by 1557 absorbance values. Asterisks represent the stepwise-selected, more informative region of spectra. M.ama: *Myrcia amazonica*; M.afa: *M.* aff. *amazonica*; M.bra: *M. bracteata*; M.cal: *M. caloneura*; M.cas: *M. castanea*; M.cre: *M. crebra*; M. cup: *M. cuprea*; M.cus: *M. cuspidata*; M.def: *M. deflexa*; M.afd: *M.* aff. *deflexa*; M.ele: *M. elevata*; M.eve: *M. eveae*; M.fas: *M. fasciculata*; M.fen: *M. fenestrata*; M.gig: *M. gigas*; M.gra: *M. grandis*; M.gui: *M. guianensis*; M.hua: *M. huallagae*; M.int: *M. intonsa*; M.lep: *M. lepida*; M.mna: *M. magna*; M.mgl: *M. magnoliifolia*; M.man: *M. manausensis*; M.nsp: *M. neospeciosa*; M.nig: *M. nigrescens*; M.oto: *M. otocalyx*; M.pai: *M.paivae*; M.pri: *M. prismatica*; M.pyr: *M. pyrifolia*; M.spl: *M. splendens*; M.syl: *M. sylvatica*; M.uau: *M. uaupensis*; M.umb: *M. umbraticola*; M.vex: *M. vexata*; M.sp1: *Myrcia* sp.1; M.sp2: *Myrcia* sp.2; M.sp3: *Myrcia* sp.3; M.sp4: *Myrcia* sp.4.

**Fig. 3.** A digitised representative leaf of each species used to extract morphometric measures for discriminant analyses. A: M.ama; B: M.afa; C: M.bra; D: M.cal; E: M.cas; F: M.cre; G: M.cup; H: M.cus; I: M.def; J: M.afd; K: M.ele; L: M.eve; M: M.fas; N: M.fen; O: M.gig; P: M.gra; Q: M.gui; R: M.hua; S: M.int; T: M.lep; U: M.mna; V: M.mgl; W: M.man; X: M.nsp; Y: M.nig; Z: M.oto; a: M.pai; b: M.pri; c: M.pyr; d: M.spl; e: M.syl; f: M.uau; g: M.umb; h: M.vex; i: M.sp1; j: M.sp2; k: M.sp3; l: M.sp4. Abbreviations of species names are presented according to Fig. 2. Scale: 1 square cm.

taxonomically useful features in the study group, we also performed analyses with randomly defined groups of specimens (i.e., a null classification model). Predictions were considered correct whenever samples were classified into the category formed by all other specimens of the same species. Correct and incorrect predictions were summarized in confusion matrices. Finally, we defined the accuracy of models as the ratio between the number of correctly predicted samples and the total number of samples. The programming routine used is presented in Supplementary File 3.

### 2.3. Species names assignment

After defining and testing local species circumscriptions (i.e., species in the context of BDFFP plots), each of them received, whenever possible, a taxonomic name derived from the literature (McVaugh, 1958; 1969; Souza et al., 1999; Holst et al., 2003), or by comparisons with herbarium specimens identified by Myrtaceae specialists (mainly Marcos Sobral and Maria Anália Duarte de Souza) or, less often, by consulting them.

### 3. Results

There were considerable variations in the spectral behaviors and leaf morphologies among the species examined (Figs. 2 and 3 respectively). The stepwise procedure indicated the wavelength region between 4,000 $cm^{-1}$ and 5,000 $cm^{-1}$ as the most informative of the NIR spectrum (Fig. 2).

Sixty-six percent of the total leaf shape variation was related to the length:width proportion (PC1, blades varying from elliptic to lanceolate), 15% was related to blades varying from ovate with an acuminate apex to obovate with an acute apex (PC2), 6% captured an elliptic-to-oblong variation (PC4), and 2% was related to blade variations from obovate with an acuminate apex to ovate with an acute apex (PC5); the 7% variation captured by PC3 and PC6 together apparently reflected
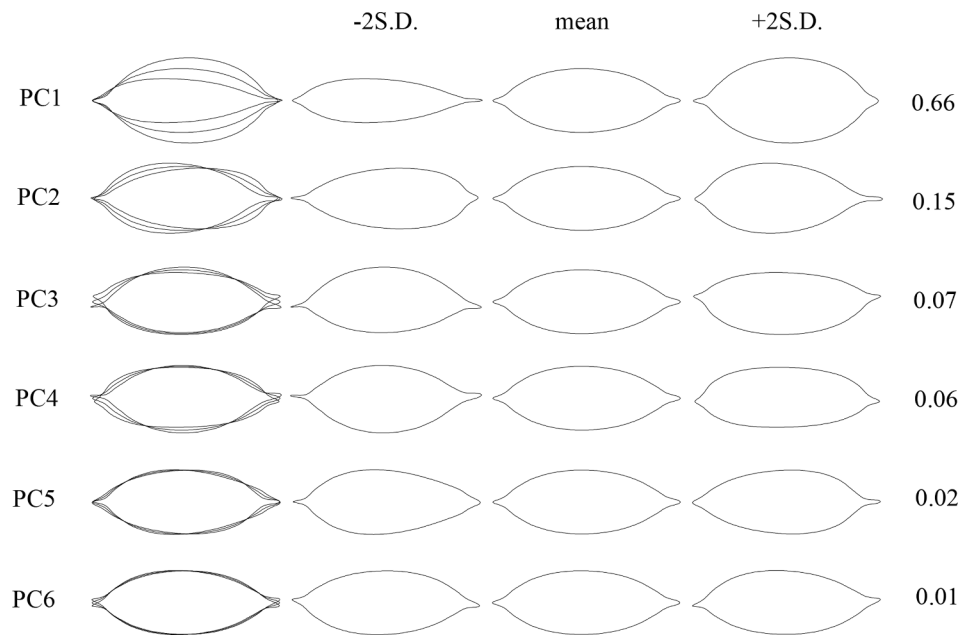
**Fig. 4.** Leaf shape variations of all species explained by the six most informative components of a Principal Component Analysis of leaf shape Fourier descriptors and their respective scores. Leaf apices are positioned on the left and petioles are placed on the right. S.D.: standard deviation.

**Table 2**
Linear Discriminant Analyses using leave-one-out cross validation. Accuracy of models built with different datasets to test the robustness of a morphotype-based classification and number of species with none, one, and more than one individual (ind.) misclassified in each model.

| Model | Accuracy | Number of species with misclassifications | | |
|---|---|---|---|---|
| | | Zero ind. | One ind. | Two or more ind. |
| Total NIRS | 98% | 31 | 5 | 1 |
| Morphometrics | 81% | 16 | 7 | 14 |
| Total NIRS + Morphometrics | 99% | 33 | 3 | 1 |
| Stepwise-selected NIRS | 97% | 29 | 6 | 2 |
| Stepwise-selected NIRS + Morphometrics | 98% | 32 | 4 | 1 |
| PCA-reduced NIRS | 97% | 29 | 6 | 2 |
| PCA-reduced NIRS + Morphometrics | 98% | 32 | 4 | 1 |

symmetry effects, which were not considered taxonomically informative and consequently disregarded in the exploratory analyses (Fig. 4).

We initially delimited 38 species of *Myrcia* among the BDFFP sites, but *Myrcia* sp.1 was represented by only a single sample and therefore could not be included in the LDA analyses. Our species hypotheses were well supported by leaf NIRS and morphometrics (accuracy > 97%) in all of the models tested, except when considering the morphometric data alone (Table 2). The combination of morphometrics with NIRS slightly increased accuracy when compared with either data set alone, reaching 99% for our classification (Fig. 5). Accuracy rates varied between 2% and 3% for random classifications. Confusion matrices between given and predicted names for all tested models are provided in Supplementary Material 4.

## 4. Discussion

Here, we show that two simple and low-cost techniques were capable of discriminating among the species of a challenging taxonomic group in a local context. The high accuracies found in the discrimination of the local species, in contrast with the lack support found for randomly generated groups, strongly indicate that the species tested are non-arbitrary phenotypic entities.

Several studies had previously demonstrated the effectiveness of NIRS data for taxonomic purposes (e.g., Lang et al., 2017; Prata et al., 2018; Damasco et al., 2019; Paiva et al., 2021), including for Myrtaceae (Castillo et al., 2008), and other workers have successfully discriminated taxa using geometric morphometrics (e.g., Esteves & Vicentini, 2013; Nery & Fiaschi, 2019; Nery et al., 2020). Spectroscopic and morphometric data require the use of a spectrophotometer and a scanner, respectively, two tools capable of processing thousands of samples after the initial expense of acquiring them, and both can be executed in repetitive work routines after only a short training period—which are desirable features for large-scale projects because of the great amount of samples to be processed. Every new census at the BDFFP, for example, entails the identification of ca. 40,000 new specimens. While the costs required for employing molecular approaches currently limit their use at massive scales (Draper et al., 2020), researchers can benefit from collecting spectroscopic and morphometric data in their attempts to better understand local diversity and speed the identification of recruits.

Highly multidimensional data should be reduced in dimension prior to analyses in order to better adjust statistical requirements (large *p* and small *n* regimes; see Nadler & Coifman, 2005). It seems, however, that the multidimensional natures of the spectra are not a serious problem when the goal is to discriminate species: we found that stepwise-selected NIRS datasets (containing 129 variables) were as efficient as the whole spectrum (containing 1557 variables) in assigning individuals to species categories. Raw NIRS variables also are highly multicollinear (Xiaobo et al., 2010), but our PCA-reduced NIRS model, containing orthogonal variables, also had a high discriminative performance. Thus, the high accuracies of the LDAs obtained in this study are more likely related to the discriminative power of data than to overfitting effects.

Although the morphometric data showed less resolving power than the spectral data, those two pieces of evidence together resulted in more powerful combined models. Two individuals of *Myrcia pyrifolia*, for instance, were misclassified as *M. magna* in the PCA-reduced NIRS model (Fig. SF4-E) and one was misclassified in both the Total NIRS and Stepwise-selected NIRS models (Figs. SF4-A and SF4-C respectively), although those two species are clearly different in terms of their overall morphologies and were split by us in the morphotyping phase. They also
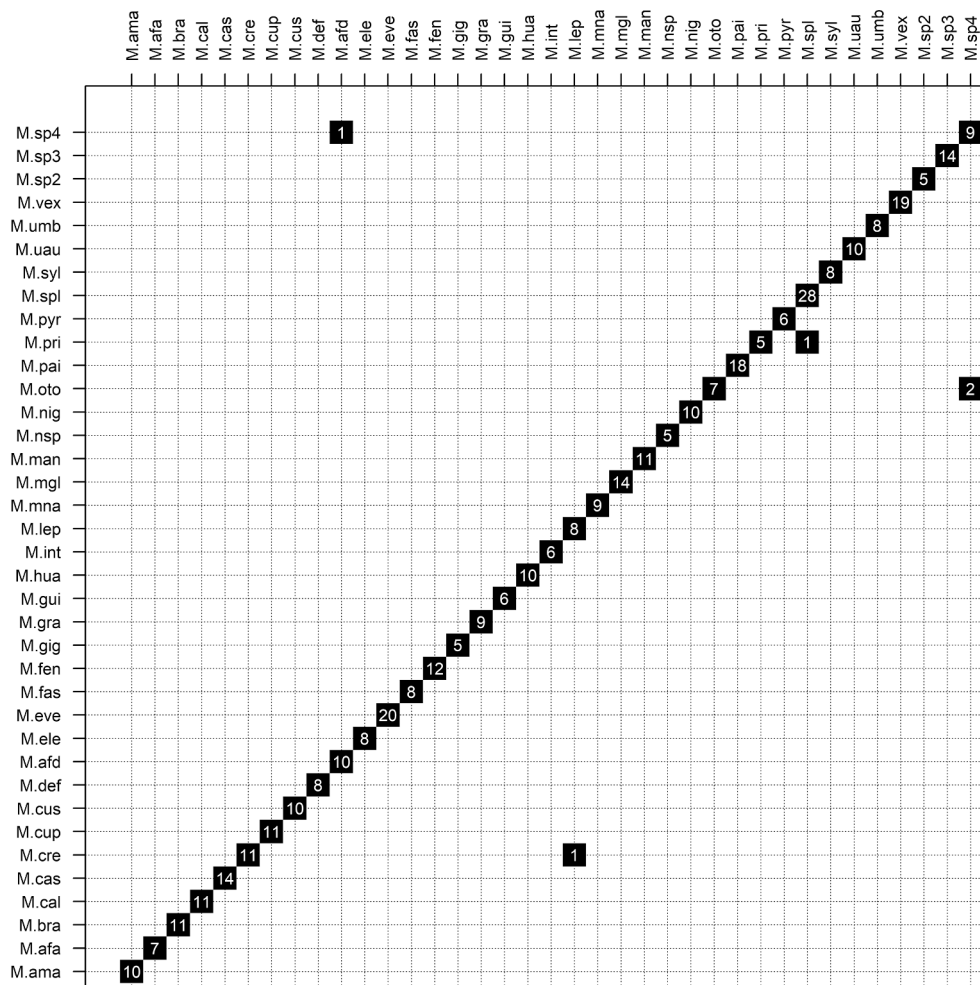
**Fig. 5.** Leave-one-out cross validation Linear Discriminant Analysis. Confusion matrix between given (rows) and predicted (columns) species identity in the Total NIRS + Morphometrics model using the morphotype-based classification. Abbreviations of species names are presented according to Fig. 2.

Confusion matrix (given row → predicted column: count):

- M.sp4 → M.afd: 1; M.sp4: 9
- M.sp3 → M.sp3: 14
- M.sp2 → M.sp2: 5
- M.vex → M.vex: 19
- M.umb → M.umb: 8
- M.uau → M.uau: 10
- M.syl → M.syl: 8
- M.spl → M.spl: 28
- M.pyr → M.pyr: 6
- M.pri → M.spl: 5; M.pri: 1
- M.pai → M.pai: 18
- M.oto → M.oto: 7; M.sp4: 2
- M.nig → M.nig: 10
- M.nsp → M.nsp: 5
- M.man → M.man: 11
- M.mgl → M.mgl: 14
- M.mna → M.mna: 9
- M.lep → M.lep: 8
- M.int → M.int: 6
- M.hua → M.hua: 10
- M.gui → M.gui: 6
- M.gra → M.gra: 9
- M.gig → M.gig: 5
- M.fen → M.fen: 12
- M.fas → M.fas: 8
- M.eve → M.eve: 20
- M.ele → M.ele: 8
- M.afd → M.afd: 10
- M.def → M.def: 8
- M.cus → M.cus: 10
- M.cup → M.cup: 11
- M.cre → M.cre: 11; M.lep: 1
- M.cas → M.cas: 14
- M.cal → M.cal: 11
- M.bra → M.bra: 11
- M.afa → M.afa: 7
- M.ama → M.ama: 10

**Table 3**

Local delimitations of species and their single corresponding accepted names according to the literature or specialists.

| BDFFP species | Accepted name | Source |
|---|---|---|
| *Myrcia amazonica*, *M.* aff. *amazonica* | *Myrcia amazonica* | (M. A. Wagner, pers. comm.) |
| *Myrcia bracteata*, *M. huallagea* | *Myrcia bracteata* | (Santos et al., 2020) |
| *Myrcia deflexa*, *M.* aff. *deflexa*, *M. eveae*, *M. magnoliifolia*, *M. otocalyx*, *M. prismatica*, *M. splendens* | *Myrcia splendens* | see Gaem et al. (2020) for a discussion about synonymy |
| *Myrcia nigrescens*, *M. umbraticola* | *Myrcia umbraticola* | (Santos et al., 2020) |

differ in terms of leaf shape and size (Fig. 3U,c) and were both well-defined in the Morphometrics model (Fig. SF4-B). It therefore seems that the 100% accurate classifications of *Myrcia pyrifolia* individuals in the models containing the two pieces of evidence (Fig. 5) were influenced by the multisource nature of the dataset.

The systematics of Neotropical Myrtaceae relies primarily on reproductive features (Landrum & Kawasaki, 1997). In general, sterile individuals *per se* did not present all of the necessary features needed to receive names in this study, although exceptions were seen with samples of *Myrcia cuprea* (O.Berg) Kiaersk. (which possess a dense coppery indumentum) and *M. magna* D.Legrand (which bear domatia on the branchlets). Four taxa represented only by sterile specimens, with no corresponding entities in herbaria nor in the literature, remained identified only to the genus level (*Myrcia* sp.1–4), but the remaining taxa could be identified based on the presence of a few fertile samples. Consequently, datasets including only a few well-curated reproductive samples can be used to build predictive models of the species identities of new sterile samples collected during inventories, especially those from tropical forests where flowering or fruiting material is rarely encountered.

Thirteen species delimited by us in the BDFFP plots belong to four species complexes, each of them usually treated under a single species name in the literature or by specialists (Table 3). While the respective synonymizations of *Myrcia huallage* McVaugh and *M. nigrescens* DC. under *M. bracteata* (Rich.) DC. and *M. umbraticola* (Kunth) E.Lucas & C. E.Wilson have been formalized (Santos et al., 2020), the close association between the morphotypes named here as *M. amazonica* DC. and *M.* aff. *amazonica* has not yet been published (M. A. Wagner, pers. comm.). The *Myrcia splendens* (Sw.) DC. Amazonian complex (MSAC) has been discussed by McVaugh (1958; 1969) and Santos (2017). In terms of this group in the BDFFP sites, Gaem et al. (2020) treat *M. eveae* Gaem & Mazine, *M. otocalyx* Gaem & L.L.Santos, and *M. prismatica* Gaem as independent species, although herbarium specimens of their individuals have previously been identified as *M. deflexa* (Poir.) DC., *M. magnoliifolia* DC., and *M. splendens* by Myrtaceae specialists. *Myrcia* aff. *deflexa*, a species to which we could not assign any formal binomial name, was recovered as independent from the other six taxa of the MSAC in this study, yet it fits the broad concept of the *M. deflexa* complex (M. Sobral,

pers. comm.). Hence, the global concepts of these species may not be accurate.

Species complexes are commonly encountered in plant systematics and demand thorough revisionary examinations (Pinheiro et al., 2018). The wide concepts applied to these entities are originated by lumping taxa together, and are favored as preliminary taxonomic solutions because of their simple definitions, and consequently easy recognition. Those concepts are often transferred to herbaria, with the result that folders of taxa belonging to species complexes frequently hold many morphologically diverse specimens collected in several localities.

The unclear morphological boundaries between taxa within species complexes are influenced, in part, by environmental variations, although environment variables tend to vary very little locally. In this context, it is not rare to encounter discrete 'morphotypes of the same variable species' in inventories, although, if there is evidence supporting their independence at any scale of coexistence, they may likely represent distinct biological species (Mayr, 1992). If we accepted the broad global morphological circumscriptions of *Myrcia amazonica*, *M. bracteata*, *M. splendens*, and *M. umbraticola* in this study regardless of our local species delimitations, the species richness of *Myrcia* in the BDFFP plots would be 24% smaller than indicated by our results (i.e., 29 taxa instead of 38). Hence, species delimitations in sympatry such as plot-based forest inventories should be treated independently of global species concepts in order to avoid underestimating local richness. Lumping similar species together into one single taxon can have direct impacts on the understanding of species abundances and distributions, as well as on downstream analyses.

Our results indicate that, when carefully conducted, local species definitions in plot-based inventories can contribute to taxonomic studies by indicating potential inconsistencies in global species circumscriptions (i.e., potential disagreements in species concepts; see De Queiroz, 2007). Such inventories must therefore be viewed by taxonomists as important sources of information. Additionally, it must be noted that herbaria may not be able to hold the increasing numbers of vouchers of fertile or sterile individuals sampled for molecular work, even though they may be fundamental to species definitions, as indicated by the few phylogenomic examinations of widespread Amazonian plant species (e.g., Prata et al. 2018, Damasco et al. 2021). Thus, the integration of data from different sources and scales is greatly needed (Draper et al. 2020).

## 5. Conclusion

This study provided successful delimitations of 38 species of a taxonomically complex group of plants that occur in permanent forest plots in the central Amazon region. We were able to show that a low-cost protocol is both practical and highly informative, and that multi-evidence perspectives are desirable in this context. Although fertile samples are fundamental for assigning names, they can be used to build predictive models containing data collected from vegetative organs that can then identify newly collected sterile samples. We showed that species delimitations before individual identifications are fundamental when studying plant assemblies, and that local species do not necessarily match global literature or herbarium concepts. Our procedure led us to understand local species richness very differently than considering morphological circumscriptions of a species throughout its entire geographical range. Such an approach may greatly affect how local community compositions are viewed and may consequently affect the results of downstream analyses in ecological studies, especially in poorly known, species-rich tropical ecosystems. Careful local species delimitations of large scale inventories may also contribute to taxonomic studies by highlighting potential inconsistencies in global definitions.

## CRediT authorship contribution statement

**Paulo Henrique Gaem:** Conceptualization, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration, Funding acquisition. **Ana Andrade:** Investigation, Resources, Supervision. **Fiorella Fernanda Mazine:** Investigation, Writing – review & editing, Supervision, Funding acquisition. **Alberto Vicentini:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Supervision.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foreco.2021.119900.

## References

Borba, E.L., Shepherd, G.J., van den Berg, C., Semir, J., 2002. Floral and vegetative morphometrics of five *Pleurothallis* (Orchidaceae) species: correlation with taxonomy, phylogeny, genetic variability and pollination systems. Ann. Bot. 90, 219–230.

Castillo, R., Contreras, D., Freer, J., Juiz, J., Valenzuela, S., 2008. Supervised pattern recognition techniques for classification of *Eucalyptus* species from leaves NIR spectra. J. Chil. Chem. Soc. 53 (4), 1709–1713.

Chapin, F.S., Reynolds, H.L., D'Antonio, C.M., Eckhart, V.M., 1996. The functional role of species in terrestrial ecosystems. In: Walker, B., Steffen, W. (Eds.), Global Change and Terrestrial Ecosystems. Cambridge University Press, Cambridge, UK, pp. 403–428.

Crawford, D.J., Stuessy, T.F., 2016. Cryptic variation, molecular data, and the challenge of conserving plant diversity in oceanic archipelagos: the critical role of plant systematics. Korean J. Plant Taxonomy 46 (2), 129–148.

Damasco, G., Baraloto, C., Vicentini, A., Daly, D.C., Baldwin, B.G., Fine, P.V.A., 2021. Revisiting the hyperdominance of Neotropical tree species under a taxonomic, functional and evolutionary perspective. Sci. Rep. 11, 9585 https://doi.org/10.1038/s41598-021-88417-y.

Damasco, G., Daly, D.C., Vicentini, A., Fine, P.A., 2019. Reestablishment of *Protium cordatum* (Burseraceae) based on integrative taxonomy. Taxon 68 (1), 34–46.

De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.

Draper, F.C., Baker, T.R., Baraloto, C., Chave, J., Costa, F., Martin, R.E., Pennington, R.T., Vicentini, A., Asner, G.P., 2020. Quantifying tropical plant diversity requires an integrated technological approach. Trends Ecol. Evol. 35 (12), 1100–1109.

Duque, A., Mueller-Landau, H.C., Valencia, R., Cardenas, D., Davies, S., Oliveira, A., Pérez, A.J., Romero-Saltos, H., Vicentini, A., 2017. Insights into regional patterns of Amazonian forest structure, biodiversity, and dominance from three large terra-firme forest dynamics plots. Biodivers. Conserv. 26, 669–686.

Durgante, F.M., Higuchi, N., Almeida, A., Vicentini, A., 2013. Species Spectral Signature: Discriminating closely related plant species in the Amazon with Near-Infrared Leaf-Spectroscopy. For. Ecol. Manage. 291, 240–248.

Esteves, S., Vicentini, A., 2013. Cryptic species in *Pagaema coriacea sensu lato* (Rubiaceae): evidence from morphology, ecology and reproductive behaviour in a sympatric context. Acta Amazonica 43 (4), 415–428.

Gaem, P.H., Santos, L.L., Andrade, A., Vicentini, A., Mazine, F.F., 2020. Three new Amazonian species of *Myrcia* sect. *Myrcia* (Myrtaceae) based on morphology and near-infrared spectroscopy. Phytotaxa 451 (4), 267–282.

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

Holst, B.K.; Landrum, L., Grifo, F. (2003). Myrtaceae. In: Berry, P.E.; Yatskievych, K. & Holst, B.K. (eds.). Flora of the Venezuelan Guayana, vol 7. Missouri Botanical Garden Press, Saint Louis, pp. 1–99.

Hopkins, M.J.G., 2007. Modelling the known and unknown plant biodiversity of the Amazon Basin. J. Biogeogr. 34, 1400–1411.

Iwata, H., Ukai, Y., 2002. SHAPE: A computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors. J. Hered. 93, 384–385.

Landrum, L.R., Kawasaki, M.L., 1997. The genera of Myrtaceae in Brazil: an illustrated synoptic treatment and identification keys. Brittonia 49, 508–536.

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

Laurance, W.F., Camargo, J.L.C., Luizão, R.C.C., Laurance, S.G., Pimm, S.L., Bruna, E.M., Stouffer, P.C., Williamson, G.B., Benítez-Malvido, J., Vasconcelos, H.L., Van Houtan, K.S., Zartman, C.E., Boyle, S.A., Didham, R.K., Andrade, A., Lovejoy, T.E., 2011. The fate of Amazonian forest fragments: a 32-year investigation. Biol. Conserv. 144, 56–67.

Lucas, E.J., Matsumoto, K., Harris, S.T., Nic Lughadha, E.M., Benardini, B., Chase, M.W., 2011. Phylogenetics, morphology, and evolution of the large genus *Myrcia* s.l. (Myrtaceae). Int. J. Plant Sci. 172 (2), 915–934.

Marenco, R.A., Camargo, M.A.B., Antezana-Vera, S.A., Oliveira, M.F., 2017. Leaf trait plasticity in six forest tree species of central Amazonia. Photosynthetica 55 (4), 679–688.

Mayr, E., 1992. A local flora and the Biological Species Concept. Am. J. Bot. 79 (2), 222–238.

McVaugh, R., 1958. Flora of Peru: Myrtaceae. Field Museum of Natural History. Botanical Series 13, 569–819.

McVaugh, R. (1969). Myrtaceae. In: Maguire, B.; Cowan, R.S., Wurdack, J.J. (eds.). The Botany of the Guayana Highland—Part VIII. Memoirs of the New York Botanical Garden 18(2): 55–286.

Nadler, B., Coifman, R.R., 2005. The prediction error in CLS and PLS: the importance of feature selection prior to multivariate calibration. J. Chemom. 19, 107–118.

Nery, E.K., Fiaschi, P., 2019. Geometric Morphometrics Dismiss the Polymorphic *Hydrocotyle quinqueloba* (Araliaceae) from the Neotropics. Syst. Bot. 44 (2), 451–469.

Nery, E.K., Matchin-Vieira, M.E., Camacho, O., Caddah, M.K., Fiaschi, P., 2020. Delimiting a constellation: integrative taxonomy of a star-shaped *Hydrocotyle* species complex (Araliaceae) from the Brazilian Atlantic forest. Plant Syst. Evol. 306, 57.

Oliveira, A.A., Mori, S., 1999. Central Amazonian terra-firme forests. I. High tree species richness on poor soils. Biodivers. Conserv. 8, 1219–1244.

Pau, G., Fuchs, F., Sklyar, O., Boutros, M., Huber, W., 2010. EBImage–an R package for image processing with applications to cellular phenotypes. Bioinformatics 26 (7), 979–981.

Pinheiro, F., Dantas-Queiroz, M.V., Palma-Silva, C., 2018. Plant species complexes as models to understand speciation and evolution: a review of South American studies. Crit. Rev. Plant Sci. 37 (1), 54–80.

Prata, E.M.B., Sass, C., Rodrigues, D.P., Domingos, F.M.C.B., Specht, C.D., Damasco, G., Ribas, C., Fine, P.V.A., Vicentini, A., 2018. Towards integrative taxonomy in Neotropical botany: disentangling the *Pagamea guianensis* species complex (Rubiaceae). Bot. J. Linn. Soc. 188, 213–231.

Paiva, D.N.A., Perdiz, R.O., Almeida, T.E., 2021. Using near-infrared spectroscopy to discriminate closely related species: a case study of neotropical ferns. J. Plant. Res. 134, 509–520.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at: <www.R-project.org>. Acessed 16 May 2020.

Santos, L.L., 2017. Taxonomia e filogenia de *Myrcia* sect. *Myrcia (Myrcia s.l.*, Myrtaceae). Universidade Federal Rural de Pernambuco, Recife, p. 319.

Santos, M.F.; Amorim, B.S.; Burton, G.P.; Fernandes, T.; Gaem, P.H.; Lourenço, A.R.L.; Lima, D.F.; Rosa, P.O.; Santos, L.L.D.; Staggemeier, V.G.; Vasconcelos, T.N.C. E Lucas, E.J. (2020). *Myrcia in* Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro. Available at: <http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB603718>. Acessed 17 May 2021.

Souza, M.A.D.; Kawasaki, M.L., Holst, B.K. (1999). Myrtaceae. In: Ribeiro, J.E.L.S.; Hopkins, M.J.G.; Vicentini, A.; Sothers, C.A.; Costa, M.A.S.; Brito, J.M.; Souza, M.A. D.; Martins, L.H.P.; Lohmann, L.G.; Assunção, P.A.C.L.; Pereira, E.C.; Silva, C.F.; Mesquita, M.R., Procópio, L.C. (eds.). Flora da Reserva Ducke: Guia de identificação das plantas vasculares de uma floresta de terra firme na Amazônia Central. INPA, Manaus, pp. 417–436.

Viscosi, V., Cardini, A., 2011. Leaf morphology, taxonomy and geometric morphometrics: a simplified protocol for beginners. PLoS ONE 6 (10), e25630.

Weihs, C., Ligges, K., Luebke, K., Raabe, N., 2005. KlaR analysing German business cycles. In: Baier, D., Decker, R., Schmidt-Thieme, L. (Eds.), Studies in Classification, Data Analysis, and Knowledge Organization: Data Analysis and Decision Support. E-Publishing Inc., Berlin, pp. 335–343.

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

Xiaobo, Z., Jiewen, Z., Povey, M.J.W., Holmes, M., Hanpin, M., 2010. Variables selection methods in near-infrared spectroscopy. Anal. Chim. Acta 667, 14–32.