Programa de Pós-Graduação em Ecologia e Conservação

Instituto de Biociências
Universidade Federal de Mato Grosso Do Sul

Molecular taxonomy as a tool for surveying the plant species diversity in the Cerrado and Pantanal

Daniel Guerra Franco

Campo Grande
Outubro 2025

Molecular taxonomy as a tool for surveying the plant species diversity in the Cerrado and Pantanal

Daniel Guerra Franco

Tese apresentada como requisito para a obtenção do título de **Doutor em Ecologia**, pelo Programa de Pós-Graduação em Ecologia e Conservação do Instituto de Biociências da Universidade Federal de Mato Grosso do Sul.

Orientadora: Aline Pedroso Lorenz
Coorientadora: Gecele Matos Paggi

Banca avaliadora

Dr. Wellington Santos Fava

Universidade Federal de Mato Grosso do Sul

Dr. Jeferson Vidart Ramos

Universidade Federal de Mato Grosso do Sul

Dr. Fabio de Oliveira Roque

Universidade Federal de Mato Grosso do Sul

Dr. Geraldo Alves Damasceno Júnior

Universidade Federal de Mato Grosso do Sul

Dr. Flávio Macedo Alves

Universidade Federal de Mato Grosso do Sul

Dra. Rachel Turba de Paula

Pontifícia Universidade Católica do Rio Grande do Sul

Dedico esse trabalho a todos que acreditaram em mim e que me apoiaram ao longo dessa campanha.

Agradecimentos

Agradeço aos meus pais Nadia e Idamir por toda dedicação à formação pessoal e profissional que tenho.

À minha companheira Bruna, minha força, meu caminho e fortaleza.

À Dona Izabel, seu Waldeir e à Brenda pelo amor admirável que dedicam à família.

Ao meu irmão Rodrigo que mesmo com a distância manteve o amor forte e sincero pela família.

À minha orientadora Aline Pedroso Lorenz e coorientadora Gecele Matos Paggi, mulheres incríveis que acreditaram em mim e nunca mediram esforços para ajudar.

A técnica do Laboratório de Ecologia e Biologia Evolutiva Josiane Vogel Cortina Theodoro e ao pesquisador Jeferson Vidart Ramos pela amizade e suporte nos dias de rotina.

Aos professores Geraldo Alves Damasceno e Flávio Macedo Alves pela ajuda nas coletas e identificação das espécies.

Ao pesquisador Wellington Santos Fava pela ajuda no uso do sequenciador de nova geração MinION.

A Fundect por me conceder a bolsa.

Ao Programa de Pós-Graduação em Ecologia e Conservação UFMS pela oportunidade e conhecimento fornecido nesses quatro anos.

Ao corpo docente da UFMS que lapidou a minha formação da graduação até aqui.

Aos meus amigos Alexei, Breno, Lenir, Renan, Rodrigo e Tadachi pelos momentos de descontração e filosofia no “Bar da Tia”.

Sumário

General Abstract

The main goal of this thesis was to show ways of achieving plant biodiversity data using the concept of DNA barcoding, and to analyse the consequences that the lack of DNA barcoding databases has for plant community surveys. In the first chapter, we developed a genomic library of tree and shrub species from an urban cerrado fragment of 36.5 hectares in Campo Grande city, Mato Grosso do Sul state. Furthermore, we have shown ways of using this library for the identification of species and the number of papers published with these species in a phylogenetic context. We were able to cover 31 species for the molecular marker *matK* and 46 species of *trnH-psbA*. We have obtained 100% of correct identification through pairwise Blast using *matK* for 6 of 16 families and 100% of correct identification using *trnH-psbA* for 4 of 19 families. The second chapter focuses on identifying which of the different plant habits of the Pantanal and which regions of the Pantanal are most represented by molecular markers submitted to GenBank. A total of five molecular markers for 2570 species were searched for availability in the public database. The representativeness was 28.4, 50, 37, 37.2, and 21 for *trnH-psbA*, *ITS*, *matK*, *rbcL*, and *trnL*, respectively, intrinsically revealing the gaps in the knowledge of Pantanal plants molecular markers. In the third chapter we have developed a barcoding library of aquatic and terrestrial plants from the Upper Paraguay River Basin in the region of Corumbá, Mato Grosso do Sul, as a tool for plant biodiversity surveys in the face of the enhancement of fire intensity in the past years. We collected a total of 93 species from 75 genus and 35 families, representing a sampling effort of 67% of the number of species described for the area. A total of 52 sequences of *trnH-psbA* and 41 of *ITS* were sequenced and deposited in Genbank. We aim to use this genomic library as a reference for surveys that focus on collecting biodiversity data using metagenomics (e.g. environmental DNA). The library is crucial

for the identification of sequences obtained in studies of plant biodiversity in the Pantanal to correctly identify the species.

Keywords: biodiversity assessment, DNA barcoding, molecular identification, conservation genetics

Resumo Geral

O principal objetivo dessa tese foi demonstrar formas de se obter dados de biodiversidade de plantas utilizando o conceito de DNA "barcoding", e analisar as consequências que a falta de bancos de dados de DNA tem para pesquisas de comunidades de plantas. No primeiro capítulo nós desenvolvemos uma biblioteca genômica de espécies arbóreas e arbustivas de em um fragmento de Cerrado urbano de 36,5 hectares em Campo Grande, Mato Grosso do Sul. Além disso, demonstramos alguns modos de usar essa biblioteca para identificação das espécies e o número de artigos publicados com as espécies da biblioteca em um contexto filogenético. Conseguimos uma cobertura de 31 espécies do marcador molecular *matK* e 46 espécies do *trnH-psbA*. Obtivemos 100% de identificações corretas com Blast pareado utilizando matK para 6 de 16 famílias e 100% de identificações corretas para 4 de 19 famílias utilizando *trnH-psbA*. O segundo capítulo foca em identificar quais dos diferentes hábitos de plantas do Pantanal e quais regiões do Pantanal estão mais representadas por marcadores moleculares submetidos no GenBank. Pesquisamos a disponibilidade de um total de 5 marcadores moleculares de 2570 espécies. A representatividade foi de 28,4, 50, 37, 37,2 e 21 para *trnH-psbA*, *ITS*, *matK*, *rbcL* e *trnL*, respectivamente, revelando intrinsecamente as lacunas de informação sobre o patrimônio genético de plantas do Pantanal. No terceiro capítulo nós desenvolvemos uma lista de DNA barcodes de plantas aquáticas e terrestres da Bacia do Alto Paraguai, na região de Corumbá, Mato Grosso do Sul, como uma das ferramentas para pesquisas de biodiversidade de plantas frente ao aumento da intensidade de fogo nos últimos anos. Foram coletadas um total de 93 espécies de 75 gêneros e 35 famílias, representando um esforço amostral de 67% do número de espécies descritas para a área. Um total de 52 sequências de *trnH-psbA* e 41 de *ITS* foram sequenciadas e depositadas no Genbank. Pretendemos usar essa biblioteca de código de barras como referência para pesquisas que focam em coletar dados de biodiversidade usando metagenômica (e.g. DNA ambiental). A biblioteca é crucial para a identificação das sequências obtidas em estudos de biodiversidade de plantas do Pantanal para a correta identificação das espécies.

Palavras-chave: avaliação da biodiversidade, DNA barcoding, identificação molecular, genética da conservação.

General Introduction

Cerrado and Pantanal, together with Caatinga, are widely recognized for areas with greater fire risk in Brazil (de Oliveira et al. 2024). Despite Cerrado and Pantanal having species with notable fire adaptations, alterations in fire regime may lead to biodiversity loss and species replacement, especially in sensitive areas, as non-savanna in Pantanal and forest vegetation in Cerrado (Schmidt et al. 2018, Leal Filho et al. 2021). Furthermore, frequently published surveys of new species described and the strong influence of collection bias suggest a lack of biodiversity data for both biomes (Oliveira et al. 2016, Colli et al. 2020). Therefore, methodologies for species identification and biodiversity assessment, such as DNA barcoding, are highly valuable in such scenarios.

Since its formalization by Hebert et al. (2003), DNA barcoding is described to be applicable widely, for example, to improve taxonomy and species identification, quantify species diversity, determine community structure and species interactions, and to protect species (Gostel & Kress 2022). This technique uses short DNA sequences (400 to 800 bp) from specimens using the difference between mean intra and interspecific genetic distance (i.e. barcoding gap) to differentiate the sequences of species. These sequences are called DNA barcodes and their efficiency may vary among taxonomic groups (Antil et al. 2023, Abdi et al. 2024). However, for good effectiveness it is important to establish a reference library that will be used for the identification of sequences, especially in cases when you urge to use the reference for metabarcoding analyses (Keck et al. 2023). The need for a reference library stimulated some projects to store and list these sequences, as does the Barcode of Life Data System (BOLD) which already hosts 14 million barcodes covering more than a million species from all over the world (Ratnasingham et al. 2024).

In contrast to DNA barcoding, metabarcoding uses bulk or environmental samples (e.g., soil, air, sediment); therefore, using an adequate set of primers, you can amplify and sequence the selected molecular marker from the genomic DNA of all species available in a sample. Both techniques have the same standard workflow: field collection, DNA extraction, amplification, sequencing, and bioinformatic analyses. The core difference is in the type of sample collected, sequencing and bioinformatic analyses (Taberlet et al. 2012, Pawlowski et al. 2020, Antil et al. 2022).

There is a description of seven classes of possible problems found in databases that may compromise species identification: mislabelling, sequencing errors, sequence conflict, taxonomic conflict, low taxonomic resolution, missing taxa, and missing intraspecific variants. The latter is caused by the high number of species represented by only one or a few sequences, as described for the molecular marker *COI* of arthropods, molluscs, and vascular plants, for instance (Weigand et al. 2019, Keck et al. 2022). Missing taxa is another issue worth of consideration. The absence of a species' sequence may promote the assignment of a sequence to a closely related species. It has been already reported that 62% of missing taxa of plants and 83% of fungi and invertebrates in Genbank (Schoch et al. 2020).

The development of a DNA barcode reference library of land and aquatic plants from Cerrado and Pantanal, and the understanding of information gaps on the public database Genbank is what pivots the three chapters of this thesis. The main goal is to use it as a complementary tool for assessment of species interactions and ocurrence. Therefore, in the first chapter we developed a local DNA barcode library using *matK* and *trnH-psbA* of tree and shrub plants from an urban Cerrado fragment in Campo Grande city, Mato Grosso do Sul state, Brazil. With this dataset, we checked the species identification accuracy of the public database GenBank and looked for how many

published papers used Cerrado plant species in a phylogenetic context and which one of these taxonomic groups are the most studied. In the second chapter, based on the survey of the 2,570 plant species described for the Pantanal, we searched the GenBank database for available sequences of five molecular markers in order to identify n order to identify gaps in the reference DNA barcode data. Finally, in the third chapter we developed a barcode library of land and aquatic plants of Upper Paraguay Basin River in the region of Corumba city using two molecular markers, *ITS* and *trnH-psbA*. We also used these markers to check the identification accuracy of GenBank database, just like we did in the first chapter, but using Pantanal land and aquatic plant sequences as queries.

Chapter 1. Assessing the accuracy of DNA barcoding for trees and shrubs identification in an urban forest reserve in the Brazilian Cerrado

Abstract

Urban forest reserves are essential strategies to guarantee the maintenance of native plant species and ecological services for urban residents. Unraveling community assembly is crucial for restoration actions, especially in a global biodiversity hotspot such as the Cerrado biome. In this work we have developed a genomic library for tree and shrub species from an urban forest reserve at the southwest region of Cerrado biome. Furthermore, we have performed a meta-data search to evaluate how many published papers used Cerrado plant species in a phylogenetic context and which taxonomic groups are the most described. We obtained 47 sequences of *matK* and 73 sequences of *trnH-psbA*, covering 31 and 46 species, respectively. Identification through pairwise Blast search using *matK* achieved 100% of correct identification for families Anacardiaceae, Combretaceae, Dileniaceae, Opiliaceae, Proteaceae, and Rubiaceae. When using *trnH-psbA* we obtained 100% of correct identification for families Bignoniaceae, Cannabaceae, Dileniaceae, and Proteaceae. The most published markers were *ITS* and *matk* with a total of 19 and 17 papers, respectively. The most published species are *Trema micranthum* and *Dipteryx alata* with a total of 12 and 8 papers, respectively. These low numbers of studies emphasize the need to characterize the genetic information of tree and shrub species from Cerrado biome to better understand community assembly providing tools for biodiversity surveys.

Keywords: reference libraries, species identification, woody plants, bibliometric analysis.

Resumo

Reservas florestais urbanas são estratégias essenciais para garantir a manutenção de espécies vegetais nativas e serviços ecológicos para moradores urbanos. Conhecer a estrutura das comunidades é crucial para ações de restauração, especialmente em um hotspot global de biodiversidade como o bioma Cerrado. Neste trabalho, desenvolvemos uma biblioteca de DNA barcode para espécies arbóreas e arbustivas de uma reserva florestal urbana na região sudoeste do bioma Cerrado. Além disso, realizamos uma busca em meta-dados para avaliar quantos artigos publicados usaram espécies de plantas do Cerrado em um contexto filogenético e quais grupos taxonômicos são os mais descritos. Obtivemos 47 sequências de *matK* e 73 sequências de *trnH-psbA*, abrangendo 31 e 46 espécies, respectivamente. A identificação por meio da busca Blast pareada usando *matK* obteve 100% de identificação correta para as famílias Anacardiaceae, Combretaceae, Dileniaceae, Opiliaceae, Proteaceae e Rubiaceae. Ao usar *trnH-psbA*, obtivemos 100% de identificação correta para as famílias Bignoniaceae, Cannabaceae, Dileniaceae e Proteaceae. Os marcadores mais publicados foram *ITS* e *matk* com um total de 19 e 17 artigos, respectivamente. As espécies mais publicadas são *Trema micranthum* e *Dipteryx alata* com um total de 12 e 8 artigos, respectivamente. Esses baixos números de estudos enfatizam a necessidade de caracterizar a informação genética de espécies arbóreas e arbustivas do bioma Cerrado para melhor compreender a estrutura da comunidade, fornecendo ferramentas para o levantamento da biodiversidade.

Palavras-chave: bibliotcas de referência, identificação de espécies, plantas lenhosas, análise bibliométrica.

1. Introduction

The growing global concern with biodiversity requires programs to map it and guide its preservation and economic exploitation. However, mapping biodiversity is hampered by the small number of specialized taxonomists and the large volume of collection and analysis work (Cognato & Caesar 2006, Pires & Marinoni 2010). Therefore, if technologies are not improved, it will take approximately 400 years to describe all neotropical flora (Azeredo 2005).

Brazil has the greatest biodiversity on the planet, with an estimated flora of around 40,000 plant species, 18,000 of which are endemic, representing approximately 10% of the world's biodiversity (Forzza et al. 2012). For this reason, Brazil's biomes are extremely important for the conservation of global biodiversity and, as they are priority areas for this purpose, they require flora surveys to support diversity inventories (Strassburg et al. 2020). However, despite two centuries of botanical exploration in the country, taxonomic knowledge is still far from being adequate for scientific and environmental needs (Thomas & Magill 2002, Sobral & Stehmann 2009).

Even today, plant species are identified based on morphological characteristics. However, in some cases, when there is a need to identify fragmented biological material, in a vegetative or juvenile state, even taxonomists may be unable to identify them. Molecular marker sequences could solve this problem, helping taxonomists to confirm identifications. For this reason, the integration of traditional taxonomy with the use of molecular markers will lead to maximum efficiency in the identification of biological material (e.g. Chase et al. 2005, Hebert & Gregory 2005, Dasmahapatra & Mallet 2006, Rubinoff et al. 2006, CBOL 2009, Ford et al. 2009, Mort et al. 2010, Hollingsworth et al. 2011).

The use of molecular markers in biological identification also plays an important role in contributing to research on taxonomy, phylogeny, and population genetics, since during the task of identifying specimens, atypical specimens may emerge that could become the target of subsequent taxonomic investigations. Furthermore, a strategic value for molecular studies with plants is the maintenance and archiving of plant tissues and extracted genomic DNA. These preserved samples provide potential raw material for research in phylogeny, population genetics, and phylogeographic studies (Hajibabaei et al. 2007). In face of this scenario, a molecular technique called DNA barcoding was developed, which aims to identify organisms at a specific level based on sequences of specific regions of DNA (i.e., Hebert et al. 2003). Therefore, the use of such markers for plant identification purposes has demonstrated high efficiency for plant identification (Kress & Erickson 2007, Ebihara et al. 2010, Kress et al. 2010, De Groot et al. 2011, Hollingsworth et al. 2011).

DNA barcodes are specific regions that can be used to identify species if there are previous sequences that can be used as a representative dataset of local diversity (Mir et al. 2021). The key characteristics of an efficient DNA barcode are (i) a short sequence (ranging from 400 to 800 bp) length to facilitate DNA amplification and sequencing; (ii) conserved flanking regions to enable primer development; and (iii) significant intra- and inter-specific variation, with a barcoding gap between (Antil et al. 2023). Due to the heterogeneity of mutation rates, a single universal DNA barcode is not optimal for plant species identification in floristic surveys. Instead, a set of standardized loci is needed to address the diversity of plant families effectively. Initially, four regions were established as standard plant DNA barcodes: *rbcL* (ribulose-1,5-bisphosphate carboxylase/oxygenase gene), *matK* (maturase K gene), and *trnH-psbA* (intergenic spacer) from chloroplast DNA (cpDNA); along with the *ITS* region (internal

transcribed spacer) from nuclear DNA (nuDNA) (Li et al. 2015, Kowalska et al. 2019, Letsiou et al. 2024). In addition to the lack of adequate variation across a wide range of families, plant DNA barcodes present other challenges. The cpDNA intergenic spacer *trnH-psbA*, for instance, shows high species discrimination, nevertheless, its application can be challenging due to significant length variation caused by poly-A/T regions and microinversions (Kim & Lee, 2005). The *ITS* region can also be challenging, as in some cases, it can exhibit significant intragenomic variation in its multiple copies across the genome (Song et al. 2012).

The Brazilian Cerrado is renowned for its biodiversity and hosts vast commodities such as cotton, corn, sugarcane, and cattle (Ferraz-Almeida & Da Mota, 2021). This biome forms a dry diagonal corridor between two other dry open vegetation biomes—Chaco in the southwest and Caatinga in the northeast—and serves as a barrier between two forest biomes, the Atlantic Forest to the southeast and the Amazon Forest to the northwest. The Cerrado contains significant headwaters that are vital water sources for most Brazilian populations (Colli et al. 2020, Klink et al. 2020). In terms of conservation, the southern and southwestern regions are less protected than the northwestern regions when considering land conversion rates and coverage by protected areas (Françoso et al. 2020). Besides, new plant species are frequently described for the Cerrado, indicating a lack of biodiversity data for this biome, which is home to many endemic and highly threatened species and is classified as a global biodiversity hotspot (Colli et al. 2020). Therefore, it is necessary to perform a bibliometric analysis for the assessment of what has already been described.

Urban forest reserves can be an effective strategy for maintaining native plant species. They also serve as refuge spots for wildlife and provide ecological services to urban residents (Yang et al. 2021). However, conserving these areas can pose

challenges when subjected to anthropogenic forces such as habitat fragmentation and biogeochemical cycles, which negatively impact plant genetic diversity and nutrient cycling, respectively (Michopoulos 2011, González et al. 2020).

This study aimed to facilitate long-term monitoring of local biodiversity by non-specialists by establishing a comprehensive DNA barcode library representing tree and shrub species from an urban Cerrado reserve. Expert botanists conducted morphological identification of the specimens, and two DNA regions were selected for sequencing. Using matK and trnH-psbA sequences and expert morphological identification, we investigated: (1) the accuracy of species identification using the National Center for Biotechnology Information (NCBI) database; (2) the most frequently published plant molecular markers; (3) the taxonomic groups with the greatest research coverage; and (4) the journals most active in publishing studies on these groups.

2. Materials and Methods

2.1.Study location

We conducted a floristic inventory in the Private Natural Heritage Reserve of the Federal University of Mato Grosso do Sul (UFMS), hereafter referred to as RPPN-UFMS, a conservation unit situated within the urban area of Campo Grande city, Mato Grosso do Sul state, Brazil (Fig. 1). In the RPPN-UFMS, there are trees typical of mesotrophic “cerradão” and deciduous forests. This mosaic composition is likely influenced by variations in soil types, with patches of fertile soil interspersed among areas with lower nutrient levels (Bueno et al. 2013). The conservation unit spans approximately 36.5 hectares (20°30’33.83” S; 54°36’57.07” W), with an average altitude of 550 meters. The climate is classified as type Aw (rainy tropical savanna),

characterized by a dry season in the winter and a rainy season in the summer (Köppen 1948). The average annual rainfall is close to 1,500 millimeters (Honer 1993).

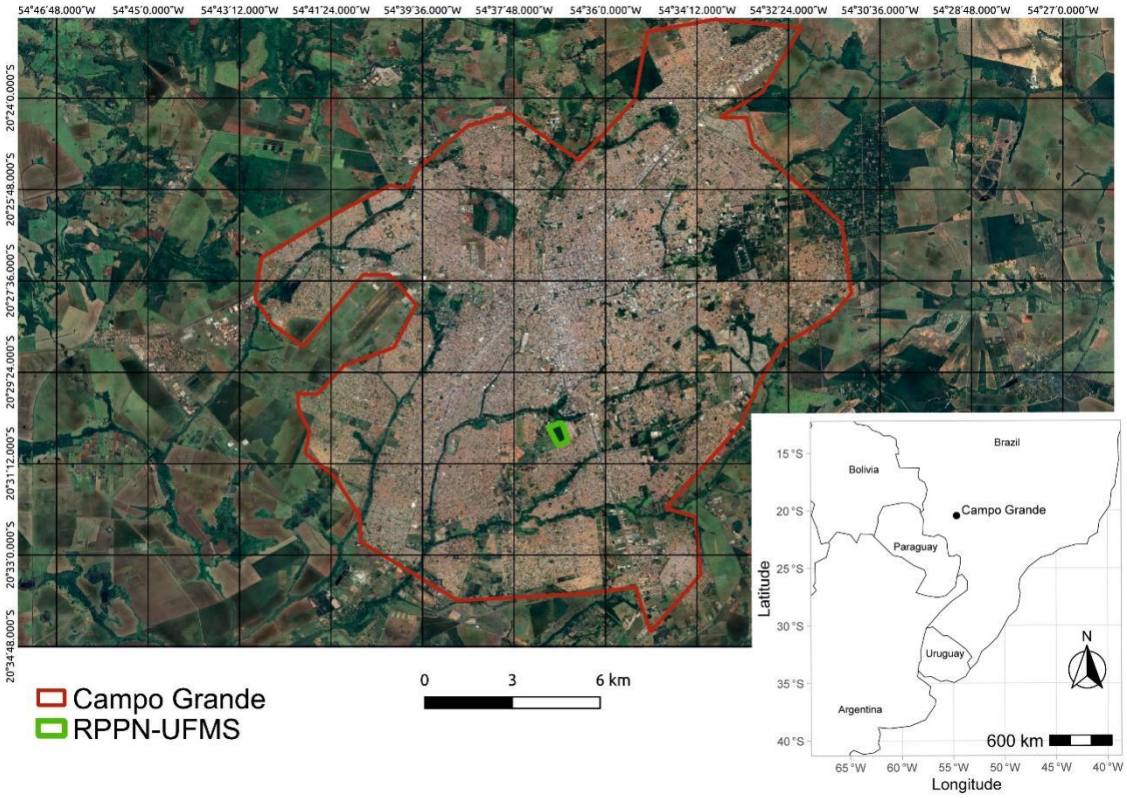

Figure 1. Map of the study area. The red polygon indicates the boundaries of the urban area of Campo Grande city, while the smaller green polygon highlights the area of the RPPN-UFMS.

2.2. Sampling

The collections took place between January 2016 and April 2018. Healthy leaves from trees and shrubs were collected and dehydrated in silica gel. An exsiccate was prepared and deposited in the CGMS herbarium for each sampled specimen. Total DNA was extracted from leaf tissue using the CTAB method, with a few modifications (Roy et al. 1992). Two DNA barcodes from the cpDNA were amplified and sequenced: *trnH-psbA,* employing the primers described by Sang et al. (1997), and *matK,* using the primers detailed by Hamilton (1999). These markers were selected due to their higher amplification success in exploratory experiments. The amplification conditions were as follows: for *trnH-psbA*, an initial denaturation at 95°C for 4 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 1 minute, and extension at 72°C for 1 minute, concluding with a final extension at 72°C for 10 minutes. For *matK*, the conditions included an initial denaturation at 94°C for 1 minute, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 1 minute, and extension at 72°C for 50 seconds, finishing with a final extension at 72°C for 5 minutes. For *trnH-psbA* the reactions contained 0.5 µM of each primer, 0.4 mM dNTPs, 1 U of Taq Polymerase (Promega), 1X PCR buffer, 2.5 mM $MgCl_2$ and approximately 20 ng of template DNA. For *matK*, each 25µL of reaction contained 1 µM of each primer, 0.2 mM dNTPs, 2 U of Taq DNA Polymerase (Promega), 1X PCR buffer, 1.5 mM $MgCl_2$ and approximately 20 ng of template DNA. Amplifications were conducted in a Mastercycler® Gradient thermocycler (Eppendorf). The products were quantified, purified with 20% PEG (polyethylene glycol), and sequenced using an ABI Prism 3100 automated DNA sequencer (Applied Biosystems). All sequences generated in this study have been submitted to GenBank and are available under accession numbers PV357259-PV357379 (Table 1).

2.3.Blast searches

Forward and reverse reads were used to generate a consensus sequence using Geneious software version 9.0. We conducted BLAST search analysis (Altschul et al. 1990) for each sample employing the blastn plugin in Geneious, retrieving the 20 top hit scores. Here, we define homonymous sequences as those with the same species named sequences available in the GenBank database. If the query sequence was homonymous with the highest hit score, it was classified as "Correct Identification". However, if among the 20 best hit scores there was at least one homonymous sequence, but it is not the highest one, it was classified as "Ambiguous Identification" (since the highest hit-score was of the same genus or family). If none of the top 20 hits scores matched the query sample's species, it was classified as "Mismatch". We also made a taxonomic level identification of sequences using the top hit score as benchmark for family, genus, or species identification of the query sequence. The BLAST search analyses considered both single-locus and complete chloroplast genome accessions. These results did not account for cases where no *trnH-psbA* or *matK* sequences of the species were present in GenBank.

2.4.Meta-analysis of GenBank Acessions

For all 49 species analyzed that we obtained at least one sequence, we searched for sequences (regardless of the genomic region) and their associated data available in GenBank. We used the scripts outlined by Maya-Lastra et al. (2022) (https://github.com/camayal/sysmexrev) as the fundamental framework to analyze the metadata of published sequences in GenBank. We utilized Datataxa (Maya-Lastra 2019) to extract information such as the paper title, author(s), journal, publication date, and the affiliated institutions. To identify similar titles (e.g., >95% similarity), we employed the Python module fuzzywuzzy (https://github.com/seatgeek/thefuzz) along with

Levenshtein distances to eliminate duplicates and then we filtered the topics using a list of terms to focus the analysis on articles of plant evolution, taxonomy, and systematics. Once we compiled the list of relevant articles (including those with no terms of evolution, taxonomy, or systematics), we manually extracted the abstracts only from published articles to incorporate into our metadata database. A thematic search in titles, abstracts, and journals was conducted using the bibtexparser module for Python (https://github.com/sciunto-org/python-bibtexparser).

3. Results

3.1. Taxonomic coverage

A total of 140 species were collected and identified in the studied area of Cerrado, which comprises 108 genera and 44 families. Of these species, 109 are trees (78%), while 31 are shrubs (22%). The richest families were Fabaceae with 30 species, Myrtaceae with seven species, Annonaceae, Malvaceae, and Vochysiaceae with six species each. We obtained 121 sequences: 46 of *matK* representing 19 families, 28 genera, and 31 species and 75 of *trnH-psbA* covering 26 families, 40 genera, and 47 species from Cerrado (Table 1). Our sampling represented about 58% of the total species reported by Bueno et al. (2013) for the RPPN-UFMS and included 14 species not previously documented. Of the 31 species for which we obtained *matK* sequences, 19.3% had no prior references in GenBank. For *trnH-psbA*, this proportion was 58.7% (Table 1). The molecular marker *matK* showed rates of correct identification, ambiguous identification, and mismatch of 44.4%, 26.8%, and 28.8%, respectively. Regarding the families, the *matK* allowed the correct identification of 100% of the Cerrado families Anacardiaceae, Combretaceae, Dileniaceae, Opiliaceae, Proteaceae, and Rubiaceae (Fig. 2a). Apocynaceae and Myrtaceae showed 100% ambiguity in

identification, while Annonaceae was 83.4%. The percentage of mismatches (e.g., unidentified) was 100% for the families Chrysobalanaceae, Ebenaceae, and Sapindaceae. Sapotaceae exhibited 33.3% ambiguity in identification and 66% mismatch sequences (Fig. 2a).

For the analyses using the marker *trnH-psbA* 27.4%, 17.4% and 55.2% of the sequences had correct identification, ambiguous identification and mismatch, respectively. We obtained 100% correct identification for species of Bignoniaceae, Cannabaceae, Dileniaceae, and Proteaceae (Fig. 2b). Ambiguous identification was encountered in species of Anacardiaceae, Clusiaceae, Fabaceae, Malpighiaceae, and Vochysiaceae, with percentage values of 100, 100, 57, 50, and 25%, respectively. Mismatches were more frequent in *trnH-psbA* compared to *matK*, with values of 100% in the families Annonaceae, Apocynaceae, Combretaceae, Ebenaceae, Erythroxylaceae, Myrsinaceae, Myrtaceae, Rubiaceae, Sapotaceae, and Siparunaceae.

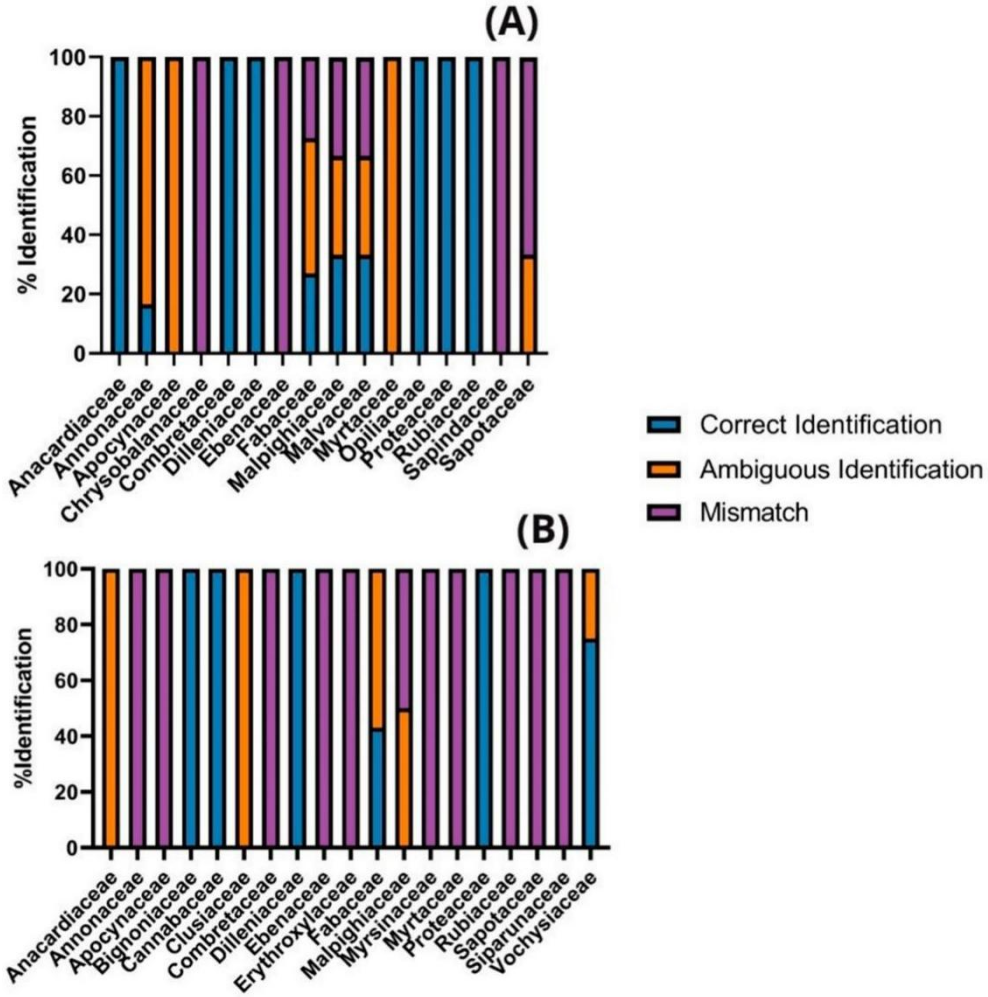

Figure 2. Identification success by family based on BLAST searches for (A) *matK* and (B) *trnH-psbA.*

Table 1. *List of tree and shrub plant species collected at the RPPN-UFMS, GenBank accession numbers and taxonomic rank of identification.*

Family	Species	Voucher CGMS	GenBank Accession Nos.		First record for this species GenBank?		Taxonomic rank with available sequences	
			matK	***trnH-psbA***	***matK***	***trnH-psbA***	***matK***	***trnH-psbA***
Anacardiaceae	*Astronium fraxinifolium* Schott	CGMS24898	PV357334	PV357259	no	no	species	genus
	Astronium urundeuva (M.Allemão) Engl.	CGMS74287	-	PV357260	-	no	-	species
	Tapirira guianensis Aubl.	CGMS24894	PV357335	-	no	-	species	-
Annonaceae	*Annona coriacea* Mart.	CGMS74323	PV357336	PV357261	no	yes	genus	genus
	Annona coriacea Mart.	CGMS24932	PV357337	PV357262	no	yes	genus	genus
	Annona crassiflora Mart.	CGMS74278	PV357338	PV357263	no	yes	genus	genus
	Annona crassiflora Mart.	CGMS24900	PV357339	PV357264	no	yes	genus	genus
	Xylopia aromatica (Lam.) Mart.	CGMS74311	PV357340	PV357265	no	no	genus	genus
	Xylopia aromatica (Lam.) Mart.	CGMS24931	PV357341	PV357266	no	no	genus	genus
	Xylopia emarginata Mart.	CGMS74387	-	PV357267	-	yes	-	genus
Apocynaceae	*Aspidosperma nobile* Müll.Arg.	CGMS74367	PV357342	-	yes	-	genus	-
	Aspidosperma tomentosum Mart. & Zucc.	CGMS74273	PV357343	-	no	-	genus	-
	Aspidosperma tomentosum Mart. & Zucc.	CGMS74277	PV357344	PV357268	no	yes	genus	genus
	Aspidosperma tomentosum Mart. & Zucc.	CGMS74322	PV357345	-	no	-	genus	-
	Himatanthus obovatus (Müll. Arg.) Woodson	CGMS74349	-	PV357269	-	yes	-	genus
	Himatanthus obovatus (Müll. Arg.) Woodson	CGMS74371	-	PV357270	-	yes	-	genus

Bignoniaceae	*Tabebuia aurea* (Silva Manso) Benth. & Hook.f. ex S.Moore	CGMS24933	-	PV357271	-	no	-	species
Bombacaceae	*Pseudobombax longiflorum* (Mart.) A.Robyns	CGMS74270	PV357346	PV357272	no	yes	genus	genus
Cannabaceae	*Trema micrantha* (L.) Blume	CGMS74365	-	PV357273	-	no	-	species
Chrysobalanaceae	*Leptobalanus humilis* Cham. & Schltdl.	CGMS24890	PV357347	PV357274	no	yes	genus	genus
	Leptobalanus humilis Cham. & Schltdl.	CGMS74324	PV357348	PV357275	no	yes	genus	genus
Clusiaceae	*Kielmeyera coriacea* Mart. & Zucc.	CGMS74340	-	PV357276	-	yes	-	genus
Combretaceae	*Terminalia corrugata* (Ducke) Gere & Boatwr.	CGMS24927	PV357349	PV357277	no*	yes	species	family
Connaraceae	*Connarus suberosus* Planch.	CGMS24893	PV357350	PV357278	yes	yes	genus	genus
Dilleniaceae	*Curatella americana* L.	CGMS74301	PV357351	PV357279	no	no	species	species
	Curatella americana L.	CGMS24926	PV357352	-	no	-	species	-
Ebenaceae	*Diospyros lasiocalyx* (Mart.) B.Walln.	CGMS74272	PV357353	PV357280	yes	yes	genus	genus
	Diospyros lasiocalyx (Mart.) B.Walln.	CGMS74336	-	PV357281	-	yes	-	genus
Erythroxylaceae	*Erythroxylum anguifugum* Mart.	CGMS74280	-	PV357282	-	yes	-	genus
	Erythroxylum anguifugum Mart.	CGMS74313	-	PV357283	-	yes	-	genus
	Erythroxylum suberosum Mart.	CGMS74296	-	PV357284	-	no*	-	genus
	Erythroxylum tortuosum Mart.	CGMS74291	-	PV357285	-	yes	-	genus
Fabaceae	*Anadenanthera peregrina* var. *falcata* (Benth.) Altschul	CGMS24905	PV357354	PV357286	no	no	genus	genus
	Anadenanthera peregrina var. *falcata* (Benth.) Altschul	CGMS74369	-	PV357287	-	no	-	genus
	Anadenanthera peregrina var. *falcata* (Benth.) Altschul	CGMS74314	PV357355	PV357288	no	no	genus	genus

	Andira cuiabensis Benth.	CGMS24902	-	PV357289	-	yes	-	genus
	Bowdichia virgilioides Kunth	CGMS74274	PV357356	PV357290	no	yes	genus	species
	Bowdichia virgilioides Kunth	CGMS74372	-	PV357291	-	yes	-	species
	Copaifera langsdorffii Desf.	CGMS74294	PV357357	PV357292	no	yes	genus	genus
	Copaifera langsdorffii Desf.	CGMS24922	PV357358	PV357293	no	yes	genus	genus
	Copaifera langsdorffii Desf.	CGMS74368	-	PV357294	-	yes	-	genus
	Dimorphandra mollis Benth	CGMS74288	PV357359	PV357295	no	no	species	species
	Dimorphandra mollis Benth	CGMS24923	-	PV357296	-	ok	-	species
	Dipteryx alata Vogel	CGMS74356	PV357360	PV357297	no	yes	genus	species
	Dipteryx alata Vogel	CGMS74312	PV357361	PV357298	no	yes	genus	species
	Diptychandra aurantiaca Tul.	CGMS74264	PV357362	PV357299	no	yes	species	species
	Hymenaea stigonocarpa Mart. ex Hayne	CGMS74353	-	PV357300	-	yes	-	genus
	Hymenaea stigonocarpa Mart. ex Hayne	CGMS74265	PV357363	PV357301	no	yes	genus	genus
	Leptolobium elegans Vogel	CGMS74343	-	PV357302	-	no	-	species
	Machaerium acutifolium Vogel	CGMS74325	-	PV357303	-	yes	-	genus
	Machaerium acutifolium Vogel	CGMS74342	-	PV357304	-	yes	-	genus
	Stryphnodendron rotundifolium Mart.	CGMS74350	-	PV357305	-	yes	-	genus
	Stryphnodendron rotundifolium Mart.	CGMS74283	-	PV357306	-	yes	-	genus
	Stryphnodendron rotundifolium Mart.	CGMS24904	-	PV357307	-	yes	-	genus
Lamiaceae	*Aegiphila integrifolia* (Jacq.) Moldenke	CGMS74344	-	PV357308	-	yes	-	genus

	Aegiphila integrifolia (Jacq.) Moldenke	CGMS74351	-	PV357309	-	yes	-	genus
Malpighiaceae	*Byrsonima coccolobifolia* Kunth	CGMS74286	PV357364	PV357310	no	no	species	genus
	Byrsonima verbascifolia (L.) DC.	CGMS74289	PV357365	PV357311	yes	no*	genus	genus
	Byrsonima verbascifolia (L.) DC.	CGMS74317	PV357366	PV357312	yes	no*	genus	genus
	Byrsonima verbascifolia (L.) DC.	CGMS74352	-	PV357313	-	no*	-	genus
Malvaceae	*Eriotheca pubescens* (Mart.) Schott & Endl.	CGMS24877	PV357367	PV357314	no	yes	species	genus
	Luehea paniculata Mart.	CGMS74284	PV357368	PV357315	no	yes	genus	genus
	Luehea paniculata Mart.	CGMS74267	PV357369	PV357316	no	yes	genus	genus
	Luehea paniculata Mart.	CGMS74354	-	PV357317	-	yes	-	genus
Myrsinaceae	*Myrsine guianensis* (Aubl.) Kuntze	CGMS24919	PV357370	PV357318	yes	no*	genus	genus
	Myrsine guianensis (Aubl.) Kuntze	CGMS74300	PV357371	PV357319	yes	no*	genus	genus
Myrtaceae	*Eugenia aurata* O.Berg	CGMS74292	PV357372	PV357320	no	no	genus	genus
Opiliaceae	*Agonandra brasiliensis* Miers ex Benth. & Hook.f.	CGMS74281	PV357373	PV357321	no	yes	species	genus
Proteaceae	*Roupala montana* Aubl.	CGMS74271	PV357374	PV357322	no	no	species	species
Rubiaceae	*Cordiera rigida* (K.Schum.) Kuntze	CGMS74315	PV357375	PV357323	no	yes	species	family
	Rudgea virburnoides (Cham.) Benth.	CGMS74276	-	PV357324	-	yes	-	genus
Sapindaceae	*Matayba guianensis* Aubl.	CGMS74334	-	PV357325	-	yes	-	genus
	Matayba guianensis Aubl.	CGMS74332	PV357376	PV357326	no	yes	genus	genus
Sapotaceae	*Chrysophyllum marginatum* (Hook. & Arn.) Radlk.	CGMS74295	PV357377	PV357327	yes	no	genus	genus
	Chrysophyllum marginatum (Hook. & Arn.) Radlk.	CGMS74363	PV357378	PV357328	yes	no	genus	genus

	Chrysophyllum marginatum (Hook. & Arn.) Radlk.	CGMS74298	PV357379	-	yes	-	genus	-
Siparunaceae	*Siparuna guianensis* Aubl.	CGMS74338	-	PV357329	-	no	-	genus
Vochysiaceae	*Qualea grandiflora* Mart.	CGMS81014	-	PV357330	-	no	-	species
	Qualea grandiflora Mart.	CGMS74333	-	PV357331	-	no	-	species
	Qualea parviflora Mart.	CGMS74310	-	PV357332	-	no	-	species
	Qualea parviflora Mart.	CGMS24914	-	PV357333	-	no	-	genus

3.2. Publication trends of the GenBank sequences

From the 136 sequences obtained in the GenBank survey, 46 were not associated with published papers and, therefore, we accounted for 90 sequences for further analysis. After applying the filter for evolution, taxonomy, and systematics, 83 sequences remained. However, we kept 90 sequences that were liked to a published paper but not necessarily had the scope of evolution, taxonomy, or systematics.

The markers most frequently used in the retrieved publications were *ITS* and matk*,* totaling 19 and 17 papers, respectively, followed by *rbcL* and *trnL-trnF* with 10 and 8 papers, respectively (Fig. 3a). Conversely, *18S*, *atpB-rbcL*, *rpl132-trnL*, *rps19*, *trnS-trnG*, and *ycf1* were used in only one publication each. *Trema micrantha* (L.) Blume (Cannabaceae) was the species with the highest number of published papers, with a total of 12 papers. *Dipteryx alata* Vogel (Fabaceae) and *Bowdichia virgilioides* Kunth (Fabaceae) have 8 and 7 published papers, respectively, while *Astronium urundeuva* (M.Allemão) Engl. (Anacardiaceae), *Dimorphandra molis* Benth. (Fabaceae), *Diptychandra aurantiaca* Tul. (Fabaceae), and *Siparuna guianensis* Aubl. (Siparunaceae) appeared in 6 papers, each. The species with the least publications were *Anadenanthera peregrine* var. *falcate* (Benth.) Altschul (Fabaceae), *Annona coriacea* Mart. (Annonaceae), *Eriotheca pubescens* (Mart.) Schott & Endl. (Malvaceae), *Eugenia aurata* O. Berg (Myrtaceae), *Hymenaea stigonocarpa* Mart. Ex Hayne (Fabaceae), *Myrsine guianensis* (Aubl.) Kuntze (Primulaceae), and *Xylopia aromatica* (Lam.) Mart. (Annonaceae) with 2 publications each; and *Aegiphila integrifolia* (Jacq.) Moldenke (Lamiaceae), *Andira cuiabensis* Benth. (Fabaceae), *Aspidosperma tomentosum* Mart. & Zucc. (Apocynaceae), *Erythroxylum anguifugum* Mart. (Erythroxylaceae), and *Erythroxylum suberosum* Mart. (Erythroxylaceae) with 1 publication each (Fig. 3b). The journals with the highest number of publications were American Journal of Botany and

Systematic Botany accounting for 12 and 7 publications, respectively, followed by Molecular Phylogenetics and Evolution and Botanical Journal of the Linnean Society with 5 publications each. We found a total of 4 publications for Taxon and Proceedings of the National Academy of Sciences, 3 for International Journal of Plant Sciences while the remaining journals had only 2 or 1 publication each (Fig. 3C).

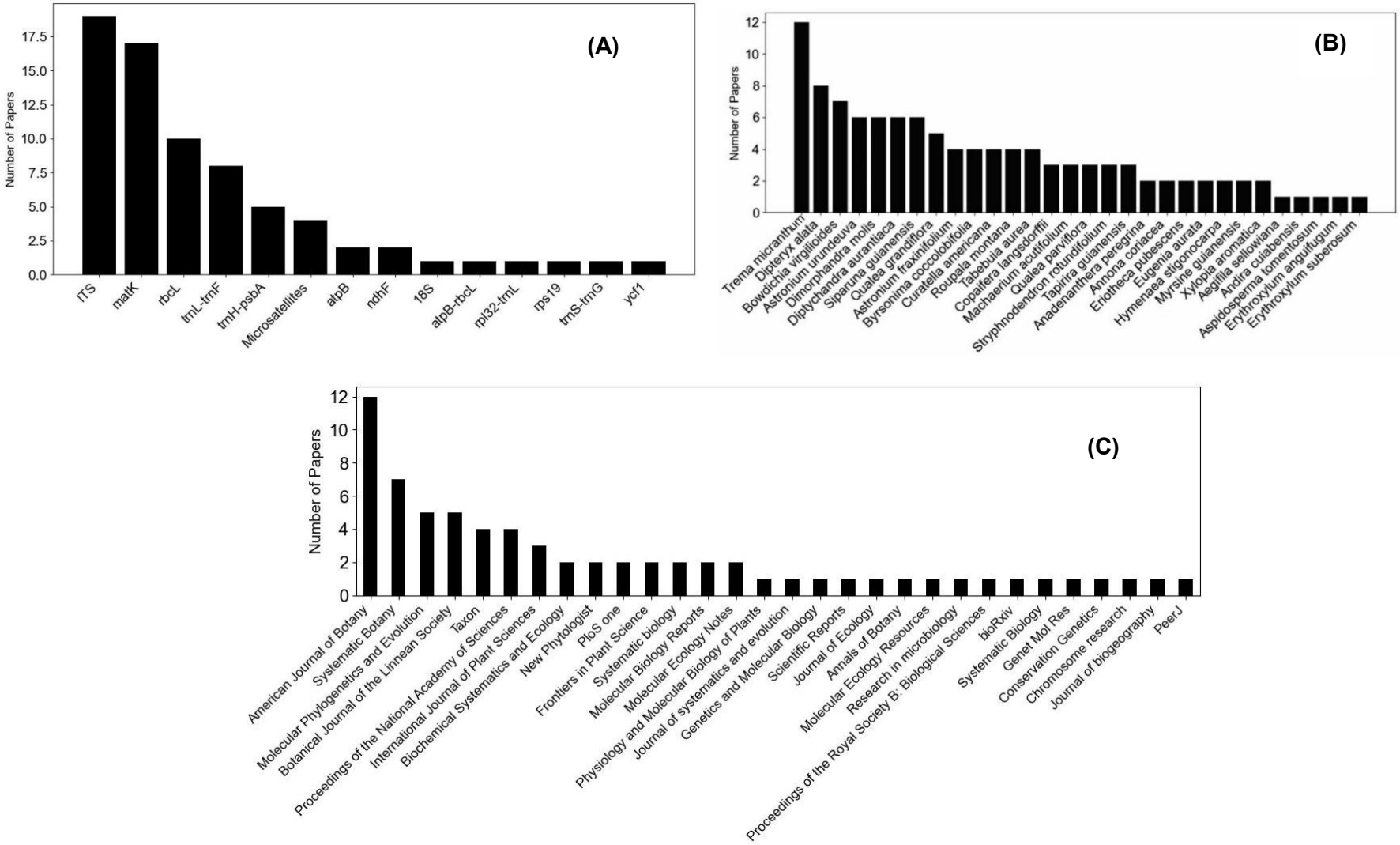

Figure 3. Meta-analysis of GenBank accession of Cerrado flora described in Table 1. Number of papers published for different (A) markers, (B) species, and (C) journals.

4. Discussion

Since the formalization of DNA barcoding as a tool for species identification, there has been a significant race to create a reference library for different taxa and regions, aiming to fill the gaps in underrepresented taxa. This effort involves multiple researchers, given the vast amount of data still waiting to be collected (Hebert et al. 2003, Joly et al. 2014). Despite the challenges of using DNA barcoding for plant identification, its benefits outweigh the limitations, as it accelerates large-scale biodiversity assessments, a crucial tool for biomes under degradation pressures, such as the Cerrado (Cowan & Fay 2012, Papadopoulou et al. 2015, Dos Santos et al. 2021). For some groups, morphological identification can be challenging and consequently more costly, especially when working with numerous samples in remote, hard-to-access areas. Furthermore, it was already reported that protected and priority areas for conservation in Cerrado also lack biodiversity surveys (Silveira et al. 2025)

Using the trees and shrubs of the RPPN-UFMS as a reference, we observed that most *trnH-psbA* sequences we obtained were recorded for the first time for these species, highlighting the necessity to enhance the DNA barcoding reference library of the Brazilian Cerrado. Similarly, Papadopoulou et al. (2015) reported a high value of underrepresented sequences of *trnH-psbA* for 437 plant species from a Nicaraguan seasonally dry forest, reaching 75% of sequences missing from GenBank. The application of DNA barcoding was already used to assess plant diversity of Amazonian "cangas", which showed that from 538 species collected, 378 had their DNA barcode described for the first time. Despite covering only 30.75% of "cangas" species using the *ITS2* barcode, when using bulk samples from Serra dos Carajás, it was possible to obtain a high taxonomic diversity in the collection sites. Generating a complete barcode library can help with restoration initiatives and the analysis of community assembly

from unexplored Amazonian areas (Vasconcelos et al. 2021). However, the composition of Amazonian plant species differ to Cerrado, highlighting the importance of the data obtained for the Cerrado plant species in this work.

When we encounter cases of ambiguous identification, such as *Astronium fraxinifolium* Schott (Anacardiaceae), there has only been one report of the *trnH-psbA* sequence submitted, dated 2007, with no publication in any journal and, therefore, no clear information about the geographic location of this specimen. Another example from our data is the *trnH-psbA* sequence of *Xylopia aromatica* (Annonaceae). The identification of this sequence was likely ambiguous for two reasons: there is only one representative sequence of *Xylopia aromatica* in GenBank, and the only available location information is that it comes from a museum in Florida. Conversely, correctly identified species, like *Tabebuia aurea* (Silva Manso) Benth. & Hook.f. ex S.Moore (Bignoniaceae) and *Qualea grandiflora* Mart. (Vochysiaceae) using *trnH-psbA*, have records of specimens from the Cerrado biome, as well as *Astronium urundeuva* which was correctly identified and have sequences of specimens from São Paulo and Bahia, Brazil.

On the other hand, *matK* has higher correct identifications and lower number of sequences recorded for the first time in GenBank, in comparison to *trnH-psbA*. We were able to achieve a higher correct identification level than a database of 113 species of plants from the Arabian Peninsula which achieved 15.04% of species level identification using *matK* (Jamdade et al. 2021). Our level of correct identification using *matK* was close to the obtained for terrestrial invasive plant species in Southwest Michigan, which reached 58.5% of 82 *matK* sequences of (Nath et al. 2024). As well as *trnH-psbA*, *matK* have applications not only for species identification but also

community identification as it was already done in water bodies in Southern Ontario, Canada (Coghlan et al. 2021).

As we have described above, markers with a lot of species' sequences reported in GenBank for the first time (i.e., low representativeness), like *trnH-psbA,* have low species identification rates, and *matK* otherwise have higher species identification rates. This must not be by chance once one of the main challenges in taxonomic reference databases are missing taxon (Keck et al. 2023). Another challenge is the lack of intraspecific variants which may have affected our species identification rate, once most of species sequences are described for the first time. Two of the species with the lowest description found in our meta-analysis are from the genus *Erythroxylum* and indeed have no correct identification in our BLAST analysis. Furthermore, none of the 9 least published species have a species level identification in our samples. On the other hand, 6 of the most described species have a species level identification for at least one marker. In addition, the use of barcoding for species identification has already been adapted for only those species which have at least one representative sequence for all the species of the genus, to avoid classification error due to lack of taxonomic resolution (West et al. 2020). Low taxonomic resolution is a problem when plenty of database vouchers are classified only at higher taxonomic levels, sometimes giving imprecise information of sequence identification (Keck et al. 2023). Another concern is that most GenBank sequences lack precise localization information, and many species are represented by only one or a few sequences, which creates a problem, as it does not account for intraspecific variation. The genetic structure of plants may be modulated even by climate changes, as have been previously described for *Anacardium occidentale*, in the Brazilian Cerrado biome (Lima et al. 2023).

Among the species' barcode sequences described for the first time, we have the first register for the *matK* of *Connarus suberosus* Planch. (Connaraceae), which have already been described as source of secondary metabolites with biological activity against *Leishmania* and Fungi (Espindola et al. 2013). For *trnH-psbA,* we have recorded the sequence of *Dipteryx alata* for the first time; this plant, known as "Baru," has food and health benefits, being commonly used in folk medicine to treat anemia, high cholesterol, diabetes, and even snake bites. The pulp is used to make cookies, liqueurs, cereal bars, and sweets, while the oil is utilized for menstrual regulation and has antipyretic and antirheumatic properties (Lima et al. 2022). Furthermore, Baru has been reported to possess potential technological uses. Its oil may be employed in pharmaceutical processes such as microencapsulation to protect active substances and transesterification for biodiesel production, respectively (Batista et al. 2012, Rojas et al. 2019). Another newly described *trnH-psbA* sequence was for *Annona crassiflora* Mart. (Annonaceae), popularly known as "Araticum", a native Cerrado tree at risk of extinction in many areas due to intensive pasture management and changes in land use (Orioli & Scariot 2021). Potential applications in the food industry have also been described for Araticum fruits, they can be used for producing sweets, jams, preserves, ice creams, cookies, cakes, juices, and liqueurs (Almeida et al. 2024). Biological, pharmacological, and nutritional activities have been attributed to Araticum, showing potential for anticholinesterase, molluscicide, insecticide, antiprotozoal, antibacterial, antiviral activities, and anticancer properties (Feitosa et al. 2017, Dev & Joseph 2021). These properties underscore the need for new tools to quickly assess Cerrado biodiversity in the context of deforestation and fire scenarios.

The sampling effort value of 57.4% of the species described by Bueno et al. (2013), includes a lot of sequences described for the first time and no previous report of

DNA barcoding library for Cerrado plants was found, suggesting this work as the first attempt to develop a DNA barcoding library for Cerrado plants. These data may be useful for the maintenance of ecological services in the RPPN-UFMS once weightening phylogenetic diversity parameters in decision making processes is necessary for biodiversity conservation efforts (Winter et al. 2013). Nevertheless, identification of species through metabarcoding analysis is directly correlated to the representativeness of the reference DNA barcode library (Alsos et al. 2018). Furthermore, it is not an easy task to achieve representativeness completeness, being already classified as one of the biggest challenges for the next decade (Kress 2017). Despite the described issues, the development of a tool for analysis of Cerrado biodiversity is useful for solving the problem of lack of species knowledge. The conscruction of a DNA barcoding library derived from an urban forest reserve may solve the species knowledge problem without the need to invest in field trips once even protected and priority areas lack biodiversity surveys (Silveira et al. 2025). There are species being described every year, intrinsically showing the lack of biodiversity information and the need for filling these gaps and reversing the conversion of Cerrado to croplands, pastures, and planted forests (Colli et al. 2020).

Chapter 2. Chasing biodiversity shortfalls: challenges in optimizing floristic surveys in the Pantanal wetlands

Abstract

This work intends to describe the current status of Pantanal plant DNA barcoding, its gaps and opportunities, and draw attention to the great lack of data and knowledge we still have about this important global biome. For this purpose, we searched for Pantanal plant species in the GenBank database to quantify the number of available species and highlight the importance of a local DNA barcode library, especially for an environment as peculiar as the Pantanal. We searched for availability in GenBank of five main molecular markers used in plants, *ITS*, *matK*, *rbcL*, *trnH-psbA*, and *trnL-trnF*) used for the identification of 2570 plant species described for the Pantanal biome, in Brazil. After surveying GenBank we found representativeness of 28.4, 50, 37, 37.2 and 21% for *trnH-psbA*, *ITS*, *matK*, *rbcL* and *trnL*, respectively. The trees were the most represented with an average representativeness of 46.6%. The herbs were the second most represented with an average of 37.1%. The parasites were the least described group with 19% representativeness, being that none of the groups reached more than 60% representativeness. In relation to the distribution of species, the exotic species and the Central West region were the most and least represented, with an average representativeness of 70.2% and 24.7%, respectively. Contrasting results like this highlight the need of a better genetic characterization of native plant species.

Keywords: biodiversity knowledge shortfalls, ecotone, molecular ecology, bioinformatics.

Resumo

Nesse trabalho levantamos a disponibilidade de sequências de DNA barcodes de plantas do Pantanal, as lacunas e oportunidades e chamamos a atenção para a grande falta de dados que ainda temos para esse bioma. Para esse propósito, pesquisamos as espécies de plantas do Pantanal no banco de dados GenBank para quantificar o número de espécies com sequências dos principais DNA barcodes disponíveis e destacar a importância de uma biblioteca de DNA barcode local, principalmente para um ambiente tão peculiar quanto o Pantanal. Buscamos a disponibilidade no GenBank de cinco principais marcadores moleculares usados em plantas, *ITS*, *matK*, *rbcL*, *trnH-psbA* e *trnL-trnF*) na identificação de 2570 espécies de plantas descritas para o bioma Pantanal, no Brasil. Após o levantamento no GenBank, encontramos a representatividade de 28,4, 50, 37, 37,2 e 21% para *trnH-psbA*, *ITS*, *matK*, *rbcL* e *trnL*, respectivamente. As árvores foram as mais representadas com média de 46,6%. As ervas foram as segundas mais representadas com 37,1% de representatividade média. Os parasitas foram o grupo menos descrito com 19% de representatividade, sendo que nenhum dos grupos atingiu mais de 60% de representatividade. Em relação à distribuição das espécies, as exóticas e a região Centro-Oeste foram as mais e menos representadas, com representatividade média de 70,2% e 24,7%, respectivamente. Resultados contrastantes como este destacam a necessidade de uma melhor caracterização genética de espécies vegetais nativas

Palavras-chave: déficit de representatividade, ecótono, ecologia molecular, bioinformática.

1. Introduction

The lack of genetic information of tropical biodiversity is concerning once there is an estimation of loss of 14,005km^2 of Pantanal native vegetation from 2018 through 2050 (Guerra et al. 2020; Kartzinel et al. 2025). The native vegetation loss is just the bottleneck to understand what is really at risk besides vegetation. An estimation of Pantanal woodland and grassland ecosystem services values reaches values of 1,588 and 2,871 US$/ha/year, respectively (Costanza et al. 1997; Bolzan et al. 2022). An evaluation of Pantanal da Nhecolandia ecosystem services was already performed and reached an annual value at US$ 15.5 billion. Some of the ecological services evaluated are gas and climate regulation, water supply and regulation and waste treatment being water supply the most valuable with an annual value at US$ 5,322 (Seidl & Moraes 2000).

Wetlands such as Pantanal are important due to their ecotone nature between terrestrial and aquatic ecosystems (Mitsch & Gosselink 2009). They have great biodiversity, and their flood cycles enhance connectivity and reduce local heterogeneity (Ward & Tockner 2001; Junk et al. 2006). Nevertheless, Pantanal has been classified as environmentally vulnerable since the 90s due to the extensive practice of agriculture land conversion, mineral transport and transformation, livestock activities, and energy generation on hydropower plants (Moretti et al. 2018).

Since its formalization in 2003, DNA barcoding has been used to identify species, mainly when there are issues regarding morphological information or their taxonomy status (Valentini et al. 2009). However, it has become a multidisciplinary subject, approaching ecology, conservation biology, medicine, pharmacology, and forensics (Joly et al. 2013; Pecnikar & Buzan 2014; Chimeno et al. 2019). The core idea is to use a determined DNA region, the so-called “barcode” to identify one or multiple

taxa, using samples such as tissue, soil, water, stool, blood, swabs, among others. The technique's advantages and popularity boosted the creation of a global consortium, the Consortium for Barcode of Life (CBOL), whose main goal is to make available at least one DNA barcode sequence for each species of the world (Yang & Zhang 2020). The development of a local DNA barcoding library makes possible the identification of new species, the construction of a community phylogeny, and the survey of unknown species interactions (Kress 2017).

Nowadays, it is possible, due to the advance of DNA barcoding techniques and analysis, to access not only an individual DNA sequence but the whole collection of DNA sequences in an environmental sample, whether it is water, stool, soil, or air, known as the metabarcoding using environmental DNA (eDNA). For example, the dietary content of vertebrates can be analyzed in a non-invasive way using stool samples (Shehzad et al. 2012; Komenova et al. 2025). Once you have previous and basic knowledge about the species' diet, you can use a specific barcode to assess the feeding habits of the individuals accurately.

The use of eDNA to assay the species composition of a sample, called "DNA metabarcoding" (Kress et al. 2014), is a study area with a particular routine pattern where you first go to the field and collect your samples, then extract, amplify, and sequence DNA barcodes in the laboratory and finally process and analyze a major data quantity at the keyboard using bioinformatics tools. Moreover, it is possible to qualify and quantify species richness and species relative abundance in the samples (Deiner et al. 2017). The development of the database is subjected to taxonomic and spatial biases. For instance, it was already reported that there are Philippines plant barcode local biodiversity is more sampled in areas with developed travel routes or near to urban areas (Berba & Matias 2022). Furthermore, to avoid identification biases and mismatches, it

is indeed imprescindible to have a thorough curation of the barcoding database. It requires the development of bioinformatics scripts such as the 'better clustering for QIIME', developed to ensure the selection of plant representative sequences according to the cluster (Banchi et al. 2020).

The main goal for a barcode is to have a significant genetic variability to distinguish species. The genetic variability of barcodes varies differently along taxonomic groups. For example, the gene that codes the cytochrome oxidase I (*COI*) in mitochondria is a good barcode for differentiation of animal species. However, when using this barcode for plants, there is not much potential for species differentiation because of its low mutation rate among plant species (Kress & Erickson 2007). Furthermore, a barcode also must have conserved regions for the development of adequate Polymerase Chain Reaction (PCR) primers. Finally, the size of a barcode is another variant that matters because it influences the sample sequencing success (CBOL 2009). Unlike barcodes for animals, for plants they are recommended to be used with concatenated loci (e.g. the combination of more than one locus for species identification). Firstly, CBOL has determined ribulose-bisphosphate carboxylase (*rbcL*) and maturase K (*matK*) coding genes as standard barcodes with supplementary use of *trnH-psbA* intergenic spacer for plants. Then, the Internal Transcribed Spacer (*ITS*) was afforded as a potential barcode by China Plant Barcode of Life (Ahmed 2022). When comparing seven different barcode markers, *ITS2* (a region from *ITS*) has been reported to have the best identification rates for medicinal plants (92.7%), for example (Chen et al. 2010). However, there are certain taxonomic groups that may demand the combination of different barcode markers for a satisfactory identification rate.

The family Lauraceae, for instance, lacks morphological differences between some species. This issue has been solved using *matK*, *ITS*, *trnH-psbA*, and *rbcL*

barcodes, with *ITS* showing the highest identification accuracy (57.5%) (Liu et al. 2017). Fabaceae, one of the largest plant families in the world, and sometimes difficult to correctly identify morphologically in dried and processed form, has also been tested for the identification with DNA barcodes. Different barcode combinations were tested and the loci with better results was *ITS2* with 40% species-level identification rate (Tahir et al. 2018). Furthermore, Apiaceae, also a flowering plant family with medical and aromatic properties has demonstrated that the combination of *ITS*, *trnH-psbA*, and *matK* as the best result for species identification, with percentage discrimination success values above 80% (Liu et al. 2014).

Nevertheless, DNA barcoding for neotropic plant communities is very scarce and for some families, like Bromeliaceae A.Juss, are a challenge once it has low sequence divergence compared to other families and therefore low identification rate, emphasizing the need for different genetic sources to improve this issue (Maia et al. 2012). A worth citing work is the development of a DNA barcode library for amazonian plants from "cangas" of the Serra dos Carajás. Barcodes for 538 species were developed, 344 of them described for the first time. They considered *ITS*2 and *rbcL* as the most suitable markers for a broad application in the region flora (Vasconcelos et al. 2021). Furthermore, in southeast Mexico, a DNA barcode library was developed for ferns and vascular plants. The group tested three plastid barcodes for vascular plants and pteridophytes, where *rbcL* showed the highest species discrimination value (66.7%) for ferns in comparison to *matK* and *trnH-psbA.* They developed new *rbcL* barcodes for 55 tropical cloud forest vascular plant species of which 48 were discriminated using maximum-likelihood phylogenetic tree approach (Trujillo-Argueta et al. 2021; Trujillo-Argueta et al. 2022).

In terms of Pantanal main plant families, Fabaceae showed accuracy level of species identification of 90% with *matK*+*rbcL* and 82.05% with *matK* (Wardani et al. 2022). Cyperaceae already had its DNA barcode developed for some species that also occur in Nigeria and in Iraq (Bello et al. 2023; Hamid et al. 2023). Poaceae, another characteristic Pantanal family, had its efficacy in DNA barcoding identification tested. The barcodes *rbcL* and *matK* had high genus identification rate (87.5-99.5%) but low species identification rate (25.6-44.6%). However, *ITS* showed high levels of genera and species identification rate and a high number of correct identification (97.6%) and low number of incorrect identifications (0.3%) (Birch et al. 2017). *ITS* indeed has a high discrimination power but its use for plant identification may have some challenges. Once DNA is usually extracted from leaves and ITS is also present in the fungal genome it is possible that the sequenced region may be from a fungus instead of the target plant. However, this problem can be solved by adding a cpDNA marker such as *trnH-psbA* (Dokane 2024).

It is important to highlight that none of the barcoding projects described so far have been developed specifically for the Pantanal flora. Furthermore, the scaffold for a good species identification using barcoding tools is the development of a DNA barcoding reference library, emphasizing the urge for such projects in Pantanal, once its local biodiversity supports ecosystem services (Grant et al. 2021; Almeida-Gomes et al. 2022). Once you have a DNA barcoding local reference library, you can use it to identify illegally logged tree species or imported and commercialized endangered plant species, just to cite a few applications. Furthermore, you can use it in crime scenes to make connections between victims, crime scenes and suspects in murder events, or in the medical area to identify poisonous plants in stomach content (Dokane 2024; Liu et al. 2021). Another benefit would be the development of new tools to study plant-animal

interactions. Conventional methods like field observation, camera-traps, biochemical, and histological analysis are usually focused on a pair of species. DNA barcoding approaches, otherwise, can analyze the interaction of multi species without the need of being in the moment of interaction indeed (Banerjee et al. 2022). However, for this analysis to be possible it is necessary to have the proper database with a minimum of representativeness.

In order to evaluate the representativeness of Pantanal plant species in GenBank we serached for a list of species for 5 DNA barcoding regions described for plants. .We expect that exotic species will be more represented once there are reports of coverage gaps in tropical biodiversity and the most represented Pantanal region will be the South and Southeast, due to its proximity to anthropic areas (Kartzinel et al. 2025). For the different habits assayed, the hypothesis is that herbs will be more represented due to its easiness for DNA extraction in comparison to trees (Tan et al. 2018).

2. Materials and Methods

2.1. *Study Area*

Pantanal is a sedimentary plain which has flood cycles caused by river or rain overflowing along the years (Silva et al. 2009). The fact of being an ecotone environment and its susceptibility to environmental pressure such as fire and flood, which promotes enhance of heterogeneity, therefore, the vegetation is classified into seven main types: forest formations, arboreal Cerrado, herbaceous Cerrado, Chaco, monodominant formations, vegetational mixtures, floristic contacts, vegetational refuges, and anthropic areas (dos Santos Vila da Silva et al. 2022). In order to contextualize the dimension of Pantanal diversity, we are going to describe the first five.

For the forest formations there are seasonal semideciduous forest, seasonal deciduous forest, and the forested savanna (known in Brazil as “Cerradão”). The forest

may occur in the margins of water courses and therefore be classified as alluvial forest or cilliary vegetation as well as sub montane forests at altitudes varying from 130 to 600 m, which resembles the Atlantic Forest from southeast Brazil and are associated with deposits of iron and manganese ores. The forested savanna is like the intermediate between forest and savanna, since it basically has the structure of forests, but the vegetation composition is mostly represented by Cerrado plant species (Ratter et al. 2003; Silva et al. 2011; dos Santos Vila da Silva et al. 2022).

The arboreal Cerrado is composed by woody savanna and park savanna. Wood savanna is characterized by discontinuous tree stratum, sandy or rocky soil, and it is amply distributed in Pantanal, mainly in areas subjected to flood. Park savanna has a physiognomy of grasslands with interspersed trees, widespread occurrencein Pantanal, especially savannas with the prevalence of *Curatella americana* or *Byrsonima cydoniifolia*. On the other hand, the herbaceous Cerrado is the grassy-woody savanna which is composed by grasslands with or without sparse or dense subshrubs or shrubs (dos Santos Vila da Silva et al. 2022).

In the southern region of Pantanal there are also vegetation formations of Chaco, comprising spiny, woody, and small species, which is similar to the Caatinga vegetation. It has two principal classifications according to the vegetation composition, the woody steppic savanna, also called open Chaco and the grassy-woody steppic savanna or Grassland. The first is composed by small trees and treelets with sparse distribution and continuous herbaceous cover. The latter is composed by herbs and grasses with small shrubs more prospect to flood pulses (Sartori et al. 2018; dos Santos Vila da Silva et al 2022).

The monodominant formations are classified according to the species dominance (a minimum of 50 or 60%) in relation to other species (Connel & Lowman 1989; Torti

et al. 2021). Furthermore, it has been related to evolutionary and ecological variables such as large seeds to overcome deep leaf litter, and shade tolerance, respectively. Pantanal may be considered the richest tropical floodplain considering the 27 different types of monodominant formations described up to now (Damasceno-Junior et al. 2022).

The annual average temperature is 24°C and is relatively constant. However, the temperature can sometimes increase to 41°C or decrease to -1°C. The temperature change is due to shifts in the air currents between the Central Brazilian Plateau in the east and the Andes in the west (Marengo et al. 2016). Annual precipitation ranges from 1,000 mm in the middle-west region to 1,600 mm along the plateaus border and the Upper Paraguay Basin mountains. Most rain occurs from October to March, with high intensity from January to March (PCBAB 1997).

2.2. DNA Barcoding Data from Pantanal Flora

We have searched five different loci, *ITS*, *trnH-psbA*, *matK*, *rbcL*, and *trnL-trnF* for the species described for Pantanal according to Damasceno-Junior & Pott (2021) in the National Center for Biotechnology Information (NCBI) nucleotide database. A total of 2.570 species were searched for each locus and for each habit and distribution. To search for the sequences of interest, we have used the Creating Reference databases for Amplicon-Based Sequencing (CRABS) package (Jeunen et al. 2009). First, we downloaded the set of loci using the filters “keyword for the locus” [All Fields] AND plants[filter]'. The “keyword for the locus” was internal transcribed spacer, *matK*, *rbcL*, *psbA* and *trnL* for each marker. After downloading, we filtered the database to obtain only sequences from 200 to 2000bp, minimum taxonomic identification to genus and remove samples from environmental DNA. Then, we used the dereplication method “single_species” to maintain only one representative sequence for each species. Finally, to search fo the Pantanal representative species, we used the subset function wich allows

to keep only the species of interest sequences. Hereafter, we say representativeness for the percentage of species sequences of determined locus available in GenBank in comparison to the species described by Damasceno-Junior & Pott (2021). Statistical analyses (one-way ANOVA and Tukey's test) were conducted in PAST v4.0 to assess differences in group representativeness (Hammer et al. 2001).

3. Results

The representativeness was 28.4, 50, 37, 37.2, and 21 for *trnH-psbA*, *ITS*, *matK*, *rbcL*, and *trnL*, respectively. In relation to the species habit, trees were the most represented with an average representativeness of 46.6% (Figure 1 F). However, considering only *trnH-psbA*, *ITS*, and *trnL*, trees were the second, fourth and second most described, with a representativeness of 39.5, 51 and 28%, respectively (Figure 1 A, B and E). The herbs were the second more represented with 37.1% of average representativeness, despite having lower values for *trnH-psbA* and *trnL-trnF* with 28.1 and 30.2%, respectively. The lianas, epiphytes, shrubs and subshrubs were the habits with intermediate values reaching 36.3, 34.9, 33.5 and 28.3% of average representativeness, respectively. The parasites were by far the least described group with an average representativeness of 19%. It was also the least described group for *trnH-psbA*, *matK*, *rbcL*, and *trnL-trnF*, but intriguingly being the most described group for *ITS*. The least described locus was *trnL-trnF* with a maximum o representativeness of 30.2% for herbs and a minimum of 5% for parasites. None of the plant habits reached more than 60% of representativeness. Despite average variation, the one-way ANOVA showed no significant difference in representativeness among growth habits ($P > 0.05$), contrary to our hypothesis of a better representation of herbs..

In relation to the distribution of the species, the most represented were the exotics with an average representativeness of 70.2%, and the group with more representativeness for all the loci separated. South, Northeast, Southeast, and North had in general similar average representativeness, 43.8, 41.6, 41.4, and 41.1, respectively (Figure 2). On the countrary, Central West demonstrated an average representativeness of 35.1% and was the least described region for *ITS*, *matK*, *rbcL*, and *trnL-trnF*. The maximum representativeness was obtained for the locus *rbcL* and exotic plants with 78.3% of representativeness and the least described was the Central West region with 24.7% of representativeness. The one-way ANOVA showed a significant difference in representativeness among Brazilian regions and exotic plants ($P < 0.05$). However, Tukey's test revealed a significant difference only in the representativeness of exotic plants compared to the other groups ($P < 0.05$), with no significant difference among Brazilian regions ($P > 0.05$)..

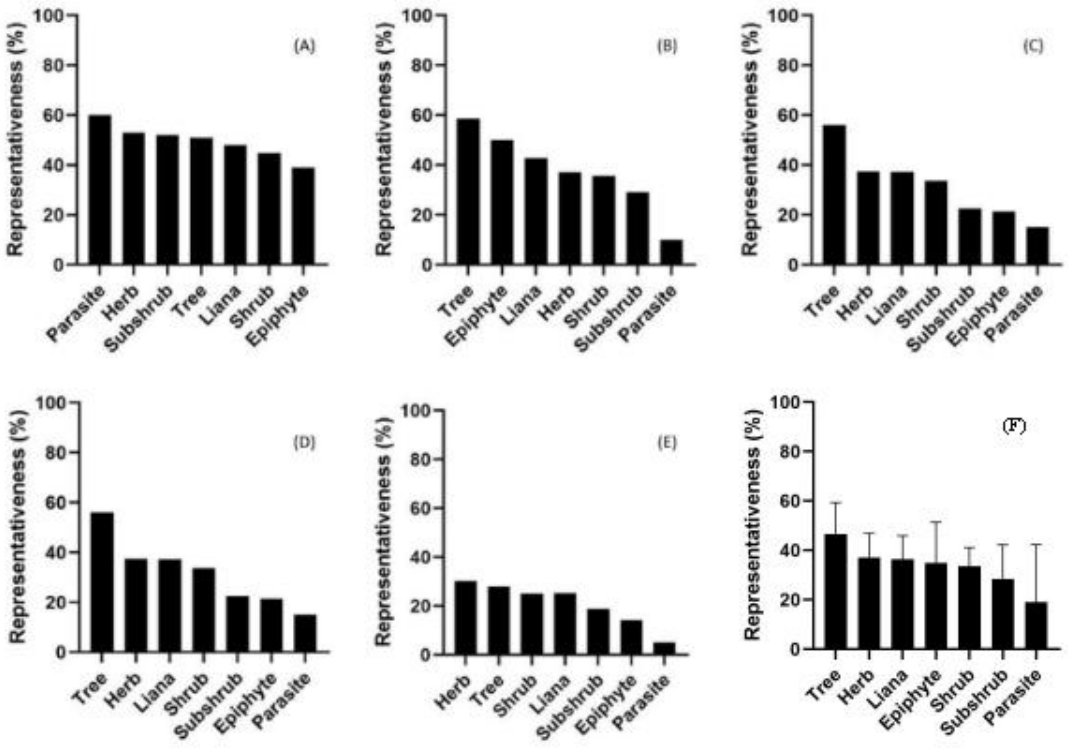

Figure 1. Representativeness of Pantanal plant species by habits for molecular markers **(A)** *trnH-psbA*, **(B)** *ITS*, **(C)**, *matK*, **(D)** *rbcL*, **(E)** *trnL-trnF*, and **(F)** Average and standard deviation of markers.

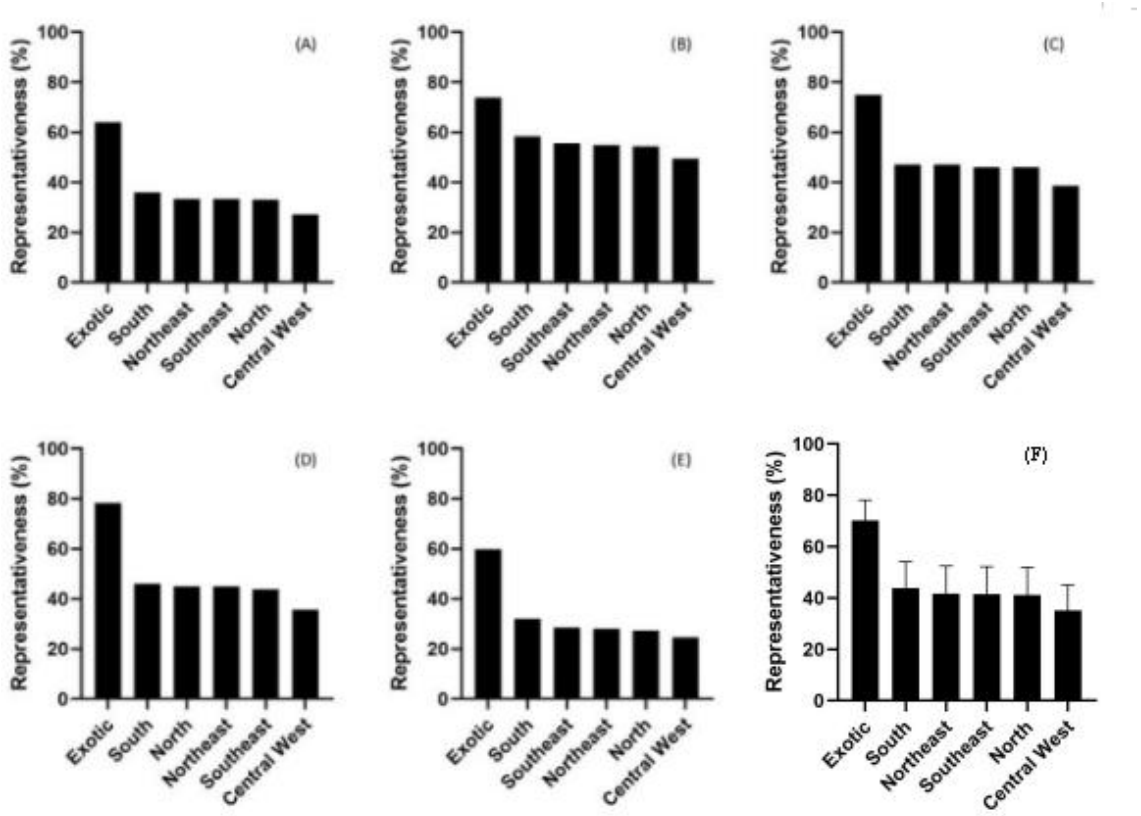

Figure 2. Representativeness of exotic and Brazilian plant species by region for molecular markers **(A)** *trnH-psbA*, **(B)** *ITS*, **(C)**, *matK*, **(D)** *rbcL*, **(E)** *trnL-trnF*, and **(F)** Average and standard deviation of markers.

4. Discussion

4.1. Extent of our ignorance

Despite recent advances, the overall coverage of plant DNA information in reference libraries for the Pantanal remains very limited. There are some representative species of Pantanal types of vegetation that have no register found for any of the loci assayed. Some worthy examples are *Terminalia argentea* Mart. & Zucc., *Terminalia corrugata* (Ducke) Gere & Boatwr., *Mouriri elliptica* Mart., *Zanthoxylum rigidum* Humb. & Bonpl. ex Willd., *Vochysia cinnamomea* Pohl, *Vochysia haenkeana* Mart., and *Vochysia rufa* Mart. from Cerrado, *Tabebuia nodosa* (Griseb.) Griseb. and *Echinopsis rhodotricha* K.Schum. from Chaco, and *Andropogon hypogynus* Hack. and *Paspalum carinatum* Humb. & Bonpl. ex Flüggé from grasslands (dos Santos Vila da Silva et al 2022). In relation to monodominant trees, *Tabebuia aurea* (Silva Manso) Benth. & Hook.f. ex S.Moore have sequences of *ITS*, *trnH-psbA*, and *rbcL*, but no sequences of *matK* and *trnL* were found. Some of the contribution come from works of dispersal dynamics of Neotropical savanna tree species (Collevatti et al. 2015). The representative species of the monodominant formation known as “Canjiqueiral”, *Byrsonima cydoniifolia* A.Juss., have sequences of *ITS* and *trnH-psbA*, but none found for *matK*, *rbcL*, and *trnL-trnF*. However, both sequences found belong to the same study, highlighting the lack of intraspecific variants in the database (de Almeida et al. 2024). For other representative species of monodominant tree and shrub formations like *Leptobalanus parvifolius* (Huber) Sothers & Prance, *Couepia uiti* (Mart. & Zucc.) Benth. ex Hook.f., and *Vochysia divergens* Pohl, there was no representative sequence found for any of the markers searched.

For the herbaceous monodominant formations, *Cyperus giganteus* Vahl have sequences of *ITS*, *matK*, and *trnH-psbA*, but no sequences of *rbcL* and *trnL-trnF*. The

records are from two different studies, which include sequences of specimens from United States, Argentina, Uruguay, and Brazil, and focus on questions of biogeography and molecular systematics (Reid et al. 2017; Pereira-Silva et al. 2025). The species of wild rice *Oryza latifolia* Desv. and *Oryza rufipogon* Griff. have representative sequences for all the markers analyzed. Indeed, there are considerable descriptions of *Oryza* loci, being already used even for securing food availability (Trinugroho et al 2023). However, we found no register of any sequences for typical representants of herbaceous monodominant formations, like *Andropogon hypogynus* Hack. (Poaceae) and *Aspilia latissima* Malme. (Asteraceae).

4.2. Implications, possibilities and strategies to overcome these gaps

The lack of representativeness in Genbank represents the absence of a protection tool for such assets. The use of eDNA has already proven to be effective for assessment of cryptic, invasive, endangered, and rare species (Liu et al. 2011; Lee et al. 2016; Hosein et al. 2017; Xu et al. 2018). In fact, when comparing traditional methods of biomonitoring with methods based on eDNA sequencing, the last has demonstrated higher resolution for all metrics of diversity tested (Gibson et al. 2015). When using the locus *trnL-trnF* amplified from soil samples for biomonitoring restrictly plants, the composition of species across varied environments matches the data obtained with classical morphological identification (Duley et al. 2023).

Even though it's demanding, this technique proves effective. It was reported a strong correlation of plant richness at family level obtained with traditional methods and with eDNA analysis, in samples all around the world. The species identification resolution was found to be more frequent in northern latitudes, highlighting the lack of data in southern latitudes (Vasar et al. 2023). Furthermore, the plant composition in meadow and heath vegetation from Varanger Peninsula was also found to match the

results obtained with traditional methods, suggesting an efficient tool for rapid assessment programs (Yoccoz et al. 2012). Both studies were performed with the P6 loop of the plastid *trnL* (UAA) intron, wich is a small sized but very informative DNA fragment (~100bp) allowing amplification even from high mois temperature and moist sites with highly degraded environmental DNA (e.g. Pantanal) (Taberlet et al. 2007; Yoccoz et al. 2012).

Efforts for conservation demand the rapid assessment of community assembly in the broad range of Pantanal sites. Despite great costs in including DNA-based methods for the development of a reference DNA barcode library, future advantages may offset the initial investment. The advantages include more cost-effective, rapid sample collection and generation of results, highly accuracy for species identification, non-invasive sampling, improvement of animal welfare, reduction of the impact of environment, and the usability for non-experts (e.g. non-professional taxonomists) (Sahu et al. 2023). For comparison purposes, studies in the northern hemisphere have already reported a higher number of plant species found using airborne eDNA metabarcoding approaches than when using traditional methods of floristics and active research (Kraaijeveld et al. 2015; Johnson et al. 2021).

4.3. *Strategies to overcome these gaps*

As mentioned, there was no significant difference between the habits and regions compared, suggesting that all groups exhibit low representativeness values. The only group that showed significantly higher representativeness compared to Brazilian native species was the exotics, highlighting the need for investment in genetic information on native plants. A possible solution is to consult the list of unavailable species for each marker (not shown here due to its length) and obtain material for DNA

extraction, amplification, sequencing, and submission to the database. Efforts should focus on priority species that currently lack representation for a given marker

In terms of political actions, Pantanal has been recognized as a National Heritage by the 1988 Brazilian Constitution, and it shelters sites of relevant international importance according to the Convention on Wetlands (https://www.ramsar.org/; Brasil 1988; Coelho-Junior et al. 2022). There are a few law reinforcements in the Brazilian federal sphere focussing on Pantanal's protection, highlighting Law n° 12.651/12 for the protection of native vegetation, Law n° 5.197/67 for fauna's protection, and Law n° 9.605/98, which provides for environmental crimes (Irigaray et al. 2017). Clearly, none of these laws focuses on the Pantanal, but consider the protection of nature and the environment. A recent state law has been approved by the state of Mato Grosso do Sul regulating specifically the Pantanal biome. The law n° 6.160 of December 18th 2023, provides for the conservation, protection, restoration, and ecologically sustainable exploration of the Restricted Use Areas of the Pantanal Plain, within the State of Mato Grosso do Sul, and creates the State Fund for the Sustainable Development of the Pantanal Biome (IMASUL 2024). The advancement of such laws and the expansion of them to federal spheres is indispensable for filling the still existing gaps of biodiversity knowledge and protecting the natural resources that Pantanal provides.

Chapter 3. DNA barcoding of plants from Upper Paraguay River Basin: a molecular approach for biodiversity assessment in the face of fire threats

Abstract

The loss of biodiversity due to extreme fire events occurring in the past years is concerning once community composition data is needed for conservation efforts. The race against time demands the development of new methods for rapid biodiversity assessment. New generation sequencing platforms has provided new strategies for community composition analysis like the use of environmental DNA, which requires a reference database for proper species level identification of sequences. In this context, we have performed floristic surveys in the Upper Paraguay River Basin, at Pantanal biome, we extracted the DNA, amplifyed and sequenced two different DNA regions: *ITS* and *trnH-psbA* from weed and shrub species, and qualify the taxonomic identification of power through BLAST search. The identification through BLAST search using *trnH-psbA* had a resolution of 28.8% of sequences at species level, 59.6% at genus, and 11.6% at family level. When using *ITS*, 46.3% of sequences had identification at species level, 41.4% at genus and 12.3% at family level. However, *ITS* and *trnH-psbA* had 41.4 and 55.7% of the sequences recorded in Genbank for the first time, suggesting a possible answer for the high rate of taxonomic identification at genus level. We highlight the gaps found in the database and possible applications if proper actions are taken.

Keywords: floristic survey, minION, internal transcribed spacer, BLAST analysis.

Resumo

A perda de biodiversidade devido a eventos de fogo extremo nos últimos anos é preocupante uma vez que dados de composição de comunidades são necessários para ações de conservação. A corrida contra o tempo demanda o desenvolvimento de novos métodos para um rápido levantamento da biodiversidade. As plataformas de sequenciamento de nova geração forneceram novas estratégias para análises de composição de comunidades, como o uso de DNA ambiental, que requer um banco de dados de referência para uma identificação das sequências a nível de espécie apropriada. Nesse contexto, nós realizamos duas florísticas na Bacia do Alto Rio Paraguai, no bioma Pantanal, nós extraímos o DNA, amplificamos e sequenciamos duas regiões de DNA diferentes: *ITS* e *trnH-psbA* de espécies de ervas e arbustos, para categorizar o poder de identificação por BLAST. A identificação por BLAST utilizando *trnH-psbA* teve uma resolução de 28,8% das sequencias a nível de espécie, 59,6% a nível de gênero e 11,6% a nível de família. Quando utilizamos *ITS*, 46,3% das sequencias tiveram uma identificação a nível de espécie, 41,4% a nível de gênero e 12,3% a nível de família. Contudo, *ITS* e *trnH-psbA* tiveram 41,4 e 55,7% das sequencias submetidas no Genbank pela primeira vez, sugerindo uma possível resposta para o alto nível de identificação a nível de gênero. Também ressaltamos as lacunas encontradas no banco de dados e as possíveis aplicações caso as medidas apropriadas forem tomadas.

Palavras-chave: levantamento florístico, minION, espaçador interno transcrito, análise BLAST.

1. Introduction

Pantanal, the largest continental wetland in the world, has been facing more severe fire regimes in the past years (Leal Filho et al. 2021). Despite a certain adaptability of Pantanal to fire regimes, if the accumulation of biomass prevails for a long time, it may promote catastrophic fire events for the flora and fauna (Damasceno-Junior et al. 2022). The biodiversity knowledge is a requirement for managers to apply conservation efforts and strategies (Rosa et al. 2021). Therefore, new methodologies for rapid and accurate plant biodiversity assessment have been developed in the past years, like DNA barcoding (Letsiou et al. 2024).

It is widely recognized that the main goal of DNA barcoding is to link an unknown sequence to a sequence of a reference database with voucher specimens. Some methods are based on comparison of sequences, recognition of short sequences and on the use of genetic distance thresholds (Altschul et al. 1990; Meier et al. 2006; Little 2011). The sequence used for identification may be nuclear like the internal transcribed spacer region (*ITS*) or chloroplastidial like the noncoding *trnH-psbA* spacer (Tnah et al. 2023). Furthermore, the next generation sequencing technologies also have provided new ways of generating biodiversity assessment through the environmental DNA metabarcoding which aims to classify a set of a species from a single sample (Deiner et al. 2017).

The eDNA samples may be derived from soil, air or water and has an applications range for a wide group of organisms, wether prokaryotes or eukaryotes (Taberlet et al. 2012; Ariza et al. 2024; Wang et al. 2024). The advantages in comparison to conventional methods include greater ability to detect close species, time and cost effectiveness and being a complementary tool (Banerjee et al. 2022). However, there are some challenges for the suitable application of this technique that includes the

development of a reference database for the identification of sequences to the minimal taxonomic level possible (Beng & Corllett 2020).

In this context, we have performed floristic surveys in the Upper Paraguay River Basin (UPRB) to extract, amplify, and sequence two different DNA regions: *ITS* and *trnH-psbA* from weed and shrub species from Pantanal. The sequences obtained were used to analyze the power of identification using the pairwise BLAST search, and we also aimed to verify which of the species obtained had the first submission of the respective sequence. The main goal was to develop a DNA barcode library for the UPRB flora as a tool for future molecular ecology studies based on environmental DNA samples.

2. Materials and Methods

2.1. Study location

Pantanal is the largest continuous tropical freshwater wetland, reaching 5,000 km² in Paraguay, 15,000 km² in Bolivia and 138,183 km² in Brazil (Silva & Abdon 1998; Damasceno-Junior & Pott 2021). We made two floristics of herbaceous, shrub, and aquatic plants in Pantanal bays of the Upper Paraguay River Basin (UPRB), in the western border of Pantanal, Corumbá city, Mato Grosso do Sul state, Brazil (Figure 1). The first collection was performed from February 28th to 8th March 2023 (wet season), and the second from 16th to 18th of August 2023 (dry season). Accordingly, to Köppen classification, the climate is seasonal Aw with mean anual temperature of 25°C with maximum values of 40°C and the rainfall occurs between November and March with 250-300 mm of precipitation and the dry season from June to August (Damasceno-Junior & Pott 2021; Martins et al. 2024). We collected only specimens with reproductive structure for morphological identification.

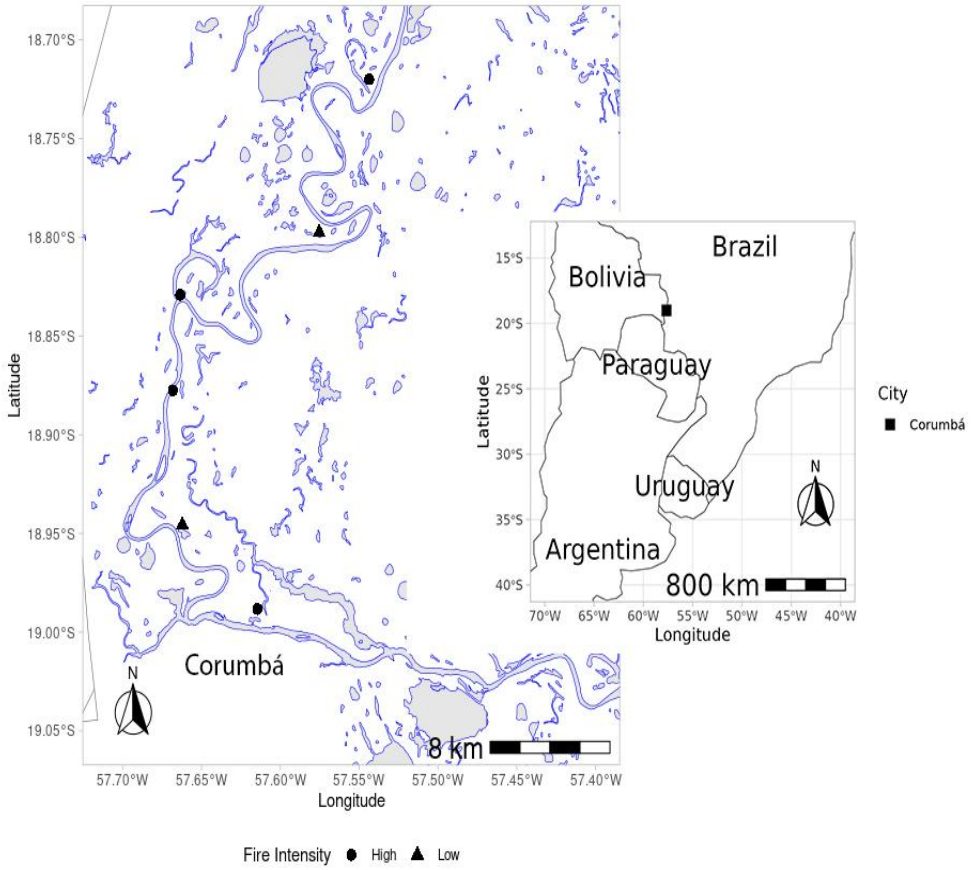

Figure 1. Collection sites in the bays of the Upper Paraguay Basin River. The square represents the city of Corumbá in a macro perspective, while the the circles and triangles represent the bays of collections of high and low fire intensity, respectively, from 1985 to 2020, according to a network named Mapbiomas (Souza et al. 2020).

2.2. Collection of data

During the first collection at dry season, the mean river height was 1.93 meters (m) while during the flood season was 4.02 m (Marinha do Brasil; https://www.marinha.mil.br/chn-6/?q=alturaAnterioresRios). For measurement of sampling effort, we searched in the virtual herbarium Reflora the number of species collected in the same area (Reflora 2023). We also collected healthy leaves and dehydrated *in silica* gel for posterior DNA extraction. The exsiccates for each specimen collected were prepared and deposited in the CGMS Herbarium. We extracted total DNA using the Wizard® Genomic DNA Purification Kit (Promega), following the manufacturer's protocol.

We used the set of primers psbA3F (5'-GTTATGCATGAACGTAATGCTC-3') and trnHR (5'-CGCGCATGGTGGATTCACAAATC-3') for amplification of *trnH-psbA* (Srirama et al. 2010) and primers ITS75 (5'-TATGCTTAAACTCAGCGGG-3') and ITS92 (5'-AAGGTTTCCGTAGGTGAAC-3') for amplification of *ITS* (Desfeux & Lejeune 1996). The PCR conditions for amplification of *trnH-psbA* included an initial denaturation at 95°C for 3 minutes, followed by 34 cycles of 95°C for 30 seconds, 55°C for 1 minute, and 72°C for 1.5 minutes, with a final denaturation at 72°C for 4 minutes. For *ITS,* the PCR started with an initial denaturation at 94°C for 5 minutes, followed by 36 cycles of 94°C for 45 seconds, 60°C for 1 minute and 72°C for 1.5 minutes, with a final denaturation at 72°C for 7 minutes. The PCR was performed with a final volume of 25µl and a final concentration of 0.3µM of each primer, 0.4µM of dNTP, 2 U of Taq DNA Polymerase (Promega), 1X PCR buffer, 3mM of $MgCl_2$, and approximately 20 ng of template DNA. The amplifications were performed in a Mastercycler® Gradient thermocycler (Eppendorf). The products were quantified and purified using a PCR products purification kit (Mebep Bioscience) and sequenced using MinION platform.

2.3. Library preparation, sequencing and bioinformatic analysis

We quantified the PCR products using a Qubit Fluorometer and the dsDNA broad-range kit (Invitrogen, Waltham, MA, USA). Then, for barcoding, 96 multiplex PCR products containing *trnH-psbA* and *ITS* amplicons were sequenced using the Rapid Barcoding Kit V14 (SQK-RBK114.24; Oxford Nanopore Technologies, Oxford, UK). We added 2.5µL of rapid barcode mix and 2.5µL of nuclease-free water into 5µL of multiplex PCR products in a concentration of approximately 100 fmol according to the manufacturer instructions. The barcode ligation was performed by incubating the mixtures at 30°C for 2 minutes followed by 80°C for 2 minutes. We loaded the DNA library on a R10.4.1 MinION flowcell (FLO-MIN114) in the platform ONT MinION MK1C (MIN-101C) and ran for 15 hours.

The basecalling and demultiplexing of the sequences were performed using the software MinKNOW with a custom analysis pipeline. We used the software Chopper v.0.8.0 to remove the reads with a quality score below 8 (Coster & Rademakers 2023). Then, we selected reference sequences in GenBank, corresponding to the molecular marker and taxon to align the reads using the software Minimap2 v.2.24-r1122 (Li 2021). Finally, after the alignment, we used the software Samtools v.1.21 to generate the consensus sequence (Danecek et al. 2021) and Geneious 7 for final typing.

2.4. Identification with pairwise Blast search

We used the pairwise BLAST search method to assay the DNA barcodes resolution for each of the loci sequences obtained for each species (Altschul et al. 1990). We used the software blastn using the Python language for BLAST searches, retrieving

the 20 best-hit scores and the respective taxon. The query sequence was classified as "Correct" if the species was homonymous to the highest hit-score. However, if among the 20 best hit-scores, but not the highest one, it was classified as "Ambiguous Identification" (since the highest hit-score was of the same genus or family). If none of the best 20 hit-scores was of the same species of query sample, it was classified as "mismatch". The taxonomic level of sequence identification was also performed using the top hit score as benchmark for family, genus, or species identification, if the species had at least one representative sequence of the respective loci in GenBank.

3. Results

3.1. Taxonomic coverage and identification through BLAST

A total of 93 species from 75 genus and 35 families were collected throughout the 2 collections. The number 93 represents a sampling effort of 67% of the number of species described for the area of collection in the Virtual Herbarium Reflora (Reflora 2023). Of these species, 26 are shrubs, 45 are herbs, and 22 are climbers. The rhichest families were Fabaceae with 17 species, Convolculaceae with 9 species, and Poaceae with 8 species. We were able to amplify and sequence a total of 52 sequences of *trnH-psbA* and 41 sequences of *ITS* (Table 1), being 1 sequence for each species and therefore 52 and 41 species for *trnH-psbA* and *ITS*, respectively. The locus *ITS* have a higher representativeness in GenBank than *trnH-psbA,* once 41.4 and 55.7% of the sequences were recorded for the first time, respectively.

The identification through BLAST search using *trnH-psbA* had a resolution of 28.8% of sequences at species level, 59.6% at genus, and 11.6% at family level. On the other hand, using *ITS*, 46.3% of sequences had identification at species level, 41.4% at genus, and 12.3% at family level, suggesting that *ITS* had a higher species identification

resolution than *trnH-psbA*. However, if we consider only the species with a previously recorded sequence in GenBank, the species identification level increases to 65% for *trnH-psbA* and 76% for *ITS*, highlighting the importance of representativeness in the database.

Table 1. List of tree and shrub plant species collected at Upper Paraguay River Basin, GenBank accession numbers, and taxonomic rank of identification. *No genus records, **only genomic records and ***considering heterotypic.

Family	Species	Voucher CGMS	GenBank Accession Nos.		First record for this species (sinonyms included) in GenBank?		Taxonomic rank with available sequences	
			trnH-psbA	*ITS*	*trnH-psbA*	*ITS*	*trnH-psbA*	*ITS*
Amaranthaceae	*Pfaffia glomerata* (Spreng.) Pedersen	CGMS88485	PV558269	PV641140	no**	no	species	family
Acanthaceae	*Justicia paludosa* (S.Moore) V.A.W.Graham	CGMS88488	PV558270	PV641141	yes	yes	genus	genus
	Justicia laevilinguis (Nees) Lindau	CGMS97022	PV558271	PV641142	yes	yes	genus	genus
Apocynaceae	*Thevetia bicornuta* Müll.Arg.	CGMS88551	PV558272	PV641143	yes	yes	genus	genus
Asteraceae	*Aspilia latissima* Malme	CGMS88519	PV558273	PV641144	yes	yes	genus***	genus***
	Enydra fluctuans Lour.	CGMS88512	PV558274	-	yes	-	genus	-
	Mikania micrantha Kunth	CGMS97020	PV558275	PV641145	no	no	species	species
	Pacourina edulis Aubl.	CGMS88487	PV558276	-	yes*	-	family	-
Boraginaceae	*Heliotropium indicum* L.	CGMS88501	PV558277	PV641146	no	no	family	species
Cabombaceae	*Cabomba furcata* Schult. & Schult.f.	CGMS88574	PV558278	PV641147	no	no	species	species
Combretaceae	*Combretum lanceolatum* Pohl ex Eichler	CGMS97019	PV558279	-	yes	-	genus	-
Convolvulaceae	*Aniseia cernua* Moric.	CGMS88577	PV558280	PV641148	yes*	yes	family	genus
	Ipomoea alba L.	CGMS88594	PV558281	PV641149	no	no	genus	species
	Ipomoea carnea Jacq.	CGMS88496	PV558282	PV641150	no	no	species	species
	Ipomoea chiliantha Hallier f.	CGMS88537	PV558283	PV641151	yes	yes	genus	genus

	Ipomoea subrevoluta Choisy	CGMS97023	PV558284	PV641152	yes	no	genus	species
	Distimake dissectus (Jacq.)A.R. Simões & Staples var. *dissectus*	CGMS88521	PV558285	PV641153	no	no	genus***	species
Cucurbitaceae	*Cayaponia podantha* Cogn.	CGMS88534	PV558286	PV641154	no	no	genus	species
	Momordica charantia L.	CGMS88490	PV558287	PV641155	no	no	species	species
Cyperaceae	*Cyperus blepharoleptos* Steud.	CGMS88580	PV558288	PV641156	no	no	species	species
Fabaceae	*Aeschynomene fluminensis* Vell.	CGMS88568	PV558289	PV641157	yes	yes	genus	genus
	Bauhinia bauhinioides (Mart.) J.F.Macbr.	CGMS88513	PV558290	-	yes	-	genus	-
	Bauhinia corniculata Benth.	CGMS88518	PV558291	PV641158	yes	yes	genus	genus
	Crotalaria incana L.	CGMS88564	PV558292	PV641159	no	no	species	species
	Macropsychanthus glaber L.P.Queiroz & Snak	CGMS97016	PV558293	PV641160	yes	yes	genus	family
	Entada polystachya (L.) DC.	CGMS88633	PV558294	PV641161	no	no	species	species
	Macroptilium atropurpureum (Sessé & Moc. ex DC.) Urb.	CGMS88570	PV558295	PV641162	no	no	genus	genus
	Neptunia plena (L.) Benth.	CGMS97028	PV558296	PV641163	yes	no	genus	species
	Senna aculeata (Pohl ex Benth.) H.S.Irwin & Barneby	CGMS88630	PV558297	-	yes	-	genus	-
	Senna alata (L.) Roxb.	CGMS88489	PV558298	PV641164	no	no	species	species
	Senna obtusifolia (L.) H.S.Irwin & Barneby	CGMS88493	PV558299	PV641165	no	no	species	species
	Senna pendula (Humb.& Bonpl.ex Willd.) H.S.Irwin & Barneby	CGMS97018	PV558300	PV641166	no	no	genus	genus
	Sesbania exasperata Kunth	CGMS88553	PV558301	PV641167	yes	no	genus	family
	Leptospron adenanthum (G.Mey.) A.Delgado	CGMS88614	PV558302	-	yes	-	genus***	-
Malpighiaceae	*Heteropterys hypericifolia* A.Juss.	CGMS97017	PV558303	-	yes	-	family	-
Malvaceae	*Byttneria filipes* Mart. ex K.Schum.	CGMS97029	PV558304	-	yes	-	genus	-
	Corchorus argutus Kunth	CGMS88587	PV558305	PV641168	yes	yes	genus	genus
	Sidastrum paniculatum (L.) Fryxell	CGMS88527	PV558306	PV641169	yes	no	genus***	species

Marsileaceae	*Marsilea crotophora* D.M.Johnson	CGMS97026	PV558307	-	yes	-	genus	-
Nymphaeaceae	*Victoria cruziana* Orb.	CGMS97030	PV558308	-	no**	-	species	-
Passifloraceae	*Passiflora gibertii N.E.Br.*	CGMS88598	PV558309	-	yes	-	genus	-
Poaceae	*Acroceras zizanioides* (Kunth) Dandy	CGMS88547	PV558310	PV641170	yes*	yes	family	genus
	Oryza latifolia Desv.	CGMS88595	PV558311	-	no	no	species	-
	Louisiella elephantipes (Nees ex Trin.) Zuloaga	CGMS97027	PV558312	PV641171	yes	yes	genus***	genus***
	Stephostachys mertensii (Rorh) Zuloaga & Morrone	CGMS88515	-	PV641172	-	yes*	-	family
	Paspalum wrightii Hitchc. & Chase	CGMS88549	PV558313	PV641173	no	yes	family	family
Pteridaceae	*Ceratopteris pteriodoides* (Hook.) Hieron.	CGMS97024	PV558314	PV641174	no**	yes	species	genus
Rubiaceae	*Borreria capitata* (Ruiz & Pav.) DC.	CGMS88603	-	PV641175	-	yes	-	genus
	Psychotria carthagenensis Jacq.	CGMS88559	PV558315	PV641176	yes	no	genus	species
Salviniaceae	*Salvinia auriculata* Aubl.	CGMS97021	PV558316	PV641177	no	no	species	genus
	Salvinia minima Baker	CGMS97025	PV558317	PV641178	no	no	species	species
Sapindaceae	*Paullinia spicata* Benth.	CGMS88545	PV558318	PV641179	no	yes	genus	genus
Solanaceae	*Solanum sisymbriifolium* Lam.	CGMS88492	PV558319	-	yes	-	genus	-
	Solanum nigrescens M.Martens & Galeotti	CGMS88498	PV558320	PV641180	yes	no	genus	species

4. Discussion

Given the number of species sequences submitted to GenBank for the first time, it is reasonable to suggest that many Pantanal plant species are poorly represented in terms of genetic information. We can therefore affirm that our efforts contribute to filling part of the genetic biodiversity gaps for Pantanal plant species. These results may provide new strategies in biomonitoring Pantanal plant species and assessing plant-animal interactions (Banerjee et al. 2022). We emphasize that without representative sequences in public database, it is impossible to conduct reliable biodiversity assessments using DNA-based approaches.

The species *Justicia laevilinguis* that had no representative sequences for any of the loci assayed is one of the species that vary on size and leaf width depending on water depth, which may promote into wrong species identification (Pott &Pott 2022). Furthermore, the identification of some species like *Thevetia bicornuta*, another with no representativeness in GenBank, may be compromised due to the short flowering period, only between January and March (Aoki et al. 2022). The database lacks information even from representative species of herbaceous monodominant formations like *Aspilia latissima* and *ITS* sequences of *Paspalum wrightii* (Damasceno-Junior et al. 2022).

The first record of sequences for these species represents the beginning of a database development but also a harsh path ahead. Plants are known to have genetic diversity within and between populations and therefore, the construction of a barcoding database demands the representativeness of more than one specimen to avoid missing intraspecific variants (Keck et al. 2022; Salgotra & Chauhan 2023). Furthermore, the lack of representative sequences of species that are source of food for birds, like *Combretum lanceolatum* would make it impossible to identify this species in a dietary study using eDNA. Other missing species like *Aeschynomene fluminensis* is foraging

source of a broad number of species such as dears, fishes and capybaras (Pott & Pott 1994).

Indeed, the use of soil eDNA for the rapid assessment of plant diversity has already been achieved in a variety of biomes like artic, boreal, tropical, and deserts (Yoccoz et al. 2012; Edwards et al. 2018; Carrasgo-Puga et al. 2021; Wang et al. 2021). The success of species identification rates has reached values of 70, 78.5 and even 100% using the noncoding *trnH-psbA* spacer through BLAST (Hoveka et al. 2016; Loera-Sánchez et al. 2020; Setsuko et al. 2023). A region of the molecular marker *ITS* named *ITS2* has already been used to assessment of plant diversity of the Amazonian canga and in the forest of Hvaler archipelago (Vasconcelos et al. 2021; Ariza et al. 2024). The barrier for the availability of eDNA as a rapid biodiversity assessment comprises not one but a set of politic, economic, and technical reasons. Those include assistance for policymaker and managers for distinguishing between eDNA research and the development of ready applications of eDNA, development of standard methods, the awareness of the utility of eDNA tools relative to traditional methods, among others (Stepien et al. 2023).

The purpose here is not only to clarify the utility and advantages of eDNA in rapid assessment of plant biodiversity but highlight the reinforcements and managements necessary for tool availability. Considering the lack of data we have found, the application of eDNA for biodiversity analysis of UPRB would face one of the main challenges of it that is linking the DNA sequences obtained to their correct taxonomic name. An achievement only possible with the availability of a reference database (Coissac et al. 2012). The efforts, however, would pay off with applications in conservation and invasion biology, biomonitoring, citizen science, and biodiversity education (Deiner et al. 2017). The advantages of richness estimates using eDNA in

comparison to traditional methods are better taxonomic resolution, higher diversity, and complementarity (Murray et al. 2012; Hawkins et al. 2015; Kraaijeveld et al. 2015). Here we say advantages but never forgetting the essentiality of traditional methods once it is crucial for correct taxonomic classification of the sequences.

General Conclusions

The data obtained so far demonstrates that the molecular markers we have analyzed, *trnH-psbA*, *matK*, *ITS* and *rbcl*, have a good discrimination power to low taxonomic levels but have some limitations due to the massive gaps that still exists in our set of data. In an ascending order, the barcodes that have the most need for representativeness are *trnL-trnF*, *trnH-psbA*, *matK, rbcL,* and *ITS*. The fact that *trnL-trnF* is one that lacks the most is concerning once it is one of the main one used for eDNA analysis. However it also reliever seeing *ITS* as the one who lacks the least needing for representativeness being also one of the most used for such studies.

As we have described through the development of this thesis, new approaches for rapid biodiversity assessment are of great importance for conservation actions. The use of environmental DNA for this purpose comprises requirements that neither the Cerrado nor the Pantanal support yet. The data described here demonstrate large plant groups that lacks more than 60% of representativeness in a popular database. We pointed the gaps and the possibilities that are being lost and somehow the neglect of initiatives for a better knowledge of our flora. On the other hand, on the perspective of a half full glass of wine, there are plenty of work ahead whether the future scenario is of urgency or not.

We believe that a worthy initiative would be the awareness of research scientists, natural resource managers, conservation policy experts and a considerable number of different government agencies, non-governmental organizations and environmental consulting firms in a collective goal for the development of tools for rapid biodiversity assessment (Stepien et al. 2023). Nevertheless, it is a long-term effort that begins with initiatives like the ones described here, addressing what the missing gaps are and how to fill them for future use. The application of molecular identification technologies in

future studies will be decisive for advancing floristic surveys as strategic tools to guide the management and conservation of plant species.

References

Abdi, G., Singh, S., Selvakumar, S., Dhar, S.K., Mudgal, G., Swaminathan, P., Tarighat, M.A., (2024). DNA Barcoding and its Applications. In Advances in Genomics: Methods and Applications (pp. 91-117). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-3169-5_5

Ahmed, S.S., (2022). DNA barcoding in plants and animals: A critical review. https://doi.org/10.20944/preprints202201.0310.v1

Almeida, R.F., Moreno, I.F., Machado, A.P.O., Meireles, M.A.A., Silva, L.K.F., Batista, E.A.C., (2024). Araticum (*Annona crassiflora* Mart.): A critical review for the food industry. Food Research International, 184:114241. https://doi.org/10.1016/j.foodres.2024.114241

Almeida-Gomes, M., de Oliveira Roque, F., Garcia, L.C., Ganci, C.C., Pacheco, E.O., Sano, N.Y., Schirpke, U., (2022). Local biodiversity supports cultural ecosystem services in the Pantanal. Wetlands, 42(7), 69. https://doi.org/10.1007/s13157-022-01579-x

Alsos, I.G., Lammers, Y., Yoccoz, N.G., Jørgensen, T., Sjögren, P., Gielly, L., Edwards, M.E., (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. PloS One, 13, e0195403. https://doi.org/10.1371/journal.pone.0195403

Altschul S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., (1990). Basic local alignment search tool. Journal of Molecular Biology, 215, 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Antil, S., Abraham, J.S., Sripoorna, S., Maurya, S., Dagar, J., Makhija, S., Bhagat, P., Gupta, R., Sood, U., Lal, R., Tojeta, R., (2023). DNA barcoding, an effective tool for species identification: a review. Molecular Biology Reports, 50, 761–775. https://doi.org/10.1007/s11033-022-08015-7

Aoki, C., Faria, R.R., Damasceno-Junior, G.A., Pott, A., (2022). Synthesis of the present knowledge on plant phenology of the Pantanal. In Flora and Vegetation of the Pantanal Wetland (pp. 535-549). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-83375-6_13

Ariza, M., Engelstad, M., Lieungh, E., Laux, M., Ready, J., Mauvisseau, Q., de Boer, H.J., (2024). Evaluating the feasibility of using plant-specific metabarcoding to assess forest types from soil eDNA. Applied Vegetation Science, 27(4), e12806. https://doi.org/10.1111/avsc.12806

Banchi, E., Ametrano, C.G., Greco, S., Stanković, D., Muggia, L., Pallavicini, A., (2020). PLANiTS: a curated sequence reference dataset for plant ITS DNA metabarcoding. Database, 2020, baz155. https://doi.org/10.1093/database/baz155

Banerjee, P., Stewart, K.A., Antognazza, C.M., Bunholi, I.V., Deiner, K., Barnes, M.A., Chen, C.Y., (2022). Plant–animal interactions in the era of environmental DNA (eDNA)—A review. Environmental DNA, 4(5), 987-999. https://doi.org/10.1002/edn3.308

Batista, A.C.F., de Souza Rodrigues, H., Pereira, N.R., Hernandez-Terrones, M.G., Vieira, A.T., Oliveira, M.F., (2012). Use of baru oil (*Dipteryx alata* Vog.) to produce biodiesel and study of the physical and chemical characteristics of biodiesel/petroleum diesel fuel blends. Chemistry and Technology of Fuels and Oils, 48, 13–16. https://doi.org/10.1007/s10553-012-0329-0

Bello, R. O., Ojokoh, A. O., Babalola, D., Shonde, T. E., Towobola, F. M. (2023). Dna Barcoding: Efficiency of Rbcl Gene for Sequence Characterisation of Cyperus Esculentus in Nigeria. https://doi.org/10.21203/rs.3.rs-2386798/v1

Beng, K.C., Corlett, R.T., (2020). Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. Biodiversity and conservation, 29(7), 2089-2121. https://doi.org/10.1007/s10531-020-01980-0

Berba, C.M.P., Matias, A.M.A., (2022). State of biodiversity documentation in the Philippines: Metadata gaps, taxonomic biases, and spatial biases in the DNA barcode data of animal and plant taxa in the context of species occurrence data. PeerJ, 10, e13146. https://doi.org/10.7717/peerj.13146

Birch, J.L., Walsh, N.G., Cantrill, D.J., Holmes, G.D., Murphy, D.J., (2017). Testing efficacy of distance and tree-based methods for DNA barcoding of grasses (Poaceae tribe Poeae) in Australia. PLoS One, 12(10), e0186259. https://doi.org/10.1371/journal.pone.0186259

Bolzan, F.P., Pereira, G.M.F., Tomas, W.M., Lourival, R., Sabino, J., Souza, F.L., de Oliveira Roque, F., (2022). Monetary value of the ecosystem services of the Pantanal and its surroundings: First approximations and perspectives. In Flora and vegetation of the Pantanal wetland (pp. 767-783). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-83375-6_21

BRASIL.[Constituição(1988)]. Constituição da República Federativa do Brasil de 1988. Brasília, DF, 05 out.1988. Available at: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. (last accessed: 15/05/2024).

Bueno, M.L., Neves, D.R.M., Souza, A.F., Oliveira Junior, E., Damasceno Junior, G.A., Pontara, V., Laura, V.A., Ratter, J.A., (2013). Influence of edaphic factors on the floristic composition of an area of cerradão in the Brazilian central-west. Acta Botanica Brasilica, 27, 445–455. https://doi.org/10.1590/S0102-33062013000200017

Carrasco-Puga, G., Díaz, F.P., Soto, D.C., Hernández-Castro, C., Contreras-López, O., Maldonado, A., Gutiérrez, R.A., (2021). Revealing hidden plant diversity in arid environments. Ecography, 44(1), 98-111. https://doi.org/10.1111/ecog.05100

CBOL Plant Working Group, 2009. A DNA barcode for land plants. Proceedings of the National academy of Sciences of the United States of America, 106, 12794–12797. https://doi.org/10.1073/pnas.0905845106

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haidar, N., Savolainen, V., 2005. Land plants and DNA barcodes: short-term and long term goals. Philosophical Transactions of the Royal Society B, 360, 1889–1895. https://doi.org/10.1098/rstb.2005.1720

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Leon, C., (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PloS one, 5(1), e8613. https://doi.org/10.1371/journal.pone.0008613

Chimeno, C., Morinière, J., Podhorna, J., Hardulak, L., Hausmann, A., Reckel, F., Haszprunar, G., (2019). DNA barcoding in forensic entomology–establishing a DNA reference library of potentially forensic relevant arthropod species. Journal of forensic sciences, 64(2), 593-601.

Coelho-Junior, M.G., Diele-Viegas, L.M., Calheiros, D.F., Silva Neto, E.C., Fearnside, P.M., Ferrante, L., (2022). Pantanal port licence would threaten the world's largest tropical wetland. Nature Ecology & Evolution, 6(5), 484-485. https://doi.org/10.1038/s41559-022-01724-z

Coghlan, S.A., Shafer, A.B., Freeland, J.R., (2021). Development of an environmental DNA metabarcoding assay for aquatic vascular plant communities. Environmental DNA, 3, 372–387. https://doi.org/10.1002/edn3.120

Cognato, A.I., Caesar, R.M., Blaxter, M.L., Vogler, A., (2006). Will DNA barcoding advance efforts to conserve biodiversity more efficiently than traditional taxonomic methods? Frontiers in Ecology and the Environment, 4, 268–273. https://doi.org/10.1890/1540-9295(2006)004[0268:WDBAET]2.0.CO;2

Coissac, E., Riaz, T., Puillandre, N., (2012). Bioinformatic challenges for DNA metabarcoding of plants and animals. Molecular ecology, 21(8), 1834-1847. https://doi.org/10.1111/j.1365-294X.2012.05550.x

Collevatti, R.G., Terribile, L.C., Rabelo, S.G., Lima-Ribeiro, M.S., (2015). Relaxed random walk model coupled with ecological niche modeling unravel the dispersal dynamics of a Neotropical savanna

tree species in the deeper Quaternary. Frontiers in Plant Science, 6, 653. https://doi.org/10.3389/fpls.2015.00653

Colli, G.R., Vieira, C.R., Dianese, J.C., (2020). Biodiversity and conservation of the Cerrado: recent advances and old challenges. Biodiversity and Conservation, 29, 1465-1475. https://doi.org/10.1007/s10531-020-01967-x.

Connell, J.H., Lowman, M.D., (1989). Low-diversity tropical rain forests: some possible mechanisms for their existence. The American Naturalist, 134(1), 88-119. https://doi.org/10.1086/284967

Costanza, R., d'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Van Den Belt, M., (1997). The value of the world's ecosystem services and natural capital. Nature, 387(6630), 253-260. https://doi.org/10.1038/387253a0

Cowan, R.S., Fay, M.F., (2012). Challenges in the DNA Barcoding of Plant Material. in: Sucher, N.J., Hennell, J.R., Carles, M.C. (Eds.), Plant DNA Fingerprinting and Barcoding. Methods in Molecular Biology. Humana Press, pp. 23–33. https://doi.org/10.1007/978-1-61779-609-8_3

Damasceno-Junior, G.A., da Silva, R.H., Gris, D., Souza, E.B.D., da Rocha, M., Manrique Pineda, D.A., Pott, A., (2022). Monodominant stands in the Pantanal. In Flora and vegetation of the pantanal wetland (pp. 393-442). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-83375-6_8

Damasceno-Junior, G.A., Pott, A. (Eds.). (2022). Flora and vegetation of the Pantanal wetland (Vol. 18). Springer Nature. https://doi.org/10.1007/978-3-030-83375-6_3

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Li, H., (2021). Twelve years of SAMtools and BCFtools. Gigascience, 10(2), giab008. https://doi.org/10.1093/gigascience/giab008

Dasmahapatra, K.K., Mallet, J., 2006. DNA barcodes: recent successes and future prospects. Heredity. 97, 254–255. https://doi.org/10.1038/sj.hdy.6800858

Da Silva, J. D. S. V., de Moura Abdon, M., Pott, A., de Arruda Mauro, R., & Sanchez, L. E. (2009). Fragile ecosystem: the Brazilian Pantanal wetland. Area Studies—Brazil: Regional Sustainable Development Review. Oxford, United Kingdom: Eolss Publishers Co. Ltd, 302-339.

De Azeredo-Espin, A.M.L., (2011). O Código de Barras da Vida baseado no DNA “Barcoding of Life”: Considerações e Perspectivas. São Paulo Press.

De Coster, W., & Rademakers, R. (2023). NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics 39: btad311. https://doi.org/10.1093/bioinformatics/btad311

De Groot, G.A., During, H.J., Mass, J.W., Schneider, H., Vogel, J.C., Erkens, R.H., (2011). Use of *rbcl* and *trnL-F* as a two-locus DNA barcode for identification of NW-European ferns: and ecological perspective. PloS One 6, e16371. https://doi.org/10.1371/journal.pone.0016371

de Oliveira Aparecido, L.E., Torsoni, G.B., Dutra, A.F., Lorençone, J.A., Lima Leite, M.R., Lorençone, P.A., de Medeiros, R.L.S., (2024). Assessing fire risk and safeguarding Brazil's biomes: a Multifactorial Approach. Theoretical and Applied Climatology, 155(9), 8815-8824. https://doi.org/10.1007/s00704-024-05163-7

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Bernatchez, L., (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular ecology, 26(21), 5872-5895. https://doi.org/10.1111/mec.14350

Desfeux, C., Lejeune, B., (1996). Systematics of Euromediterranean *Silene* (Caryophyllaceae): evidence from a phylogenetic analysis using ITS sequences. Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie, 319(4), 351-358.

Dev, A.R.A., Joseph, S.M., (2021). Anticancer potential of *Annona* genus: A detailed review. Journal of the Indian Chemical Society, 98, 100231. https://doi.org/10.1016/j.jics.2021.100231

Dos Santos, G.L., Pereira, M.G., Delgado, R.C., Magistrali, I.C., da Silva, C.G., de Oliveira, C.M.M., Larangeira, J.P.B., da Silva, T.P., (2021). Degradation of the Brazilian Cerrado: Interactions with human disturbance and environmental variables. Forest Ecology and Management,. 482, 118875. https://doi.org/10.1016/j.foreco.2020.118875

Dokane, K., (2024). Analysis of markers for forensic plant species identification. Forensic Science International, 358, 112007. https://doi.org/10.1016/j.forsciint.2024.112007

dos Santos Vila da Silva, J., Pott, A., Chaves, J.V.B. (2021). Classification and Mapping of the Vegetation of the Brazilian Pantanal. In: Damasceno-Junior, G.A., Pott, A. (eds) Flora and Vegetation of the Pantanal Wetland. Plant and Vegetation, vol 18. Springer, Cham. https://doi.org/10.1007/978-3-030-83375-6_2

Duley, E., Iribar, A., Bisson, C., Chave, J., Donald, J., (2023). Soil environmental DNA metabarcoding can quantify local plant diversity for biomonitoring across varied environments. Restoration Ecology, 31(4), e13831. https://doi.org/10.1111/rec.13831

Ebihara, A., Nitta, J.H., Ito, M., (2010). Molecular species identification with rich floristic sampling: DNA barcoding the pteridophyte flora of Japan. PloS One 5, e15136. https://doi.org/10.1371/journal.pone.0015136

Edwards, M.E., Alsos, I.G., Yoccoz, N., Coissac, E., Goslar, T., Gielly, L., Taberlet, P., (2018). Metabarcoding of modern soil DNA gives a highly local vegetation signal in Svalbard tundra. The Holocene, 28(12), 2006-2016. https://doi.org/10.1177/0959683618798095

Espindola, L.S., Costa, R., Santana, D., de Paula, J., Araújo, R., Braz-Filho R., (2013). Investigation of Secondary Metabolites in Plants of the Brazilian Cerrado Against Leishmania and Fungi. Planta Medica, 79, CL7. https://doi.org/10.1055/s-0033-1348532

Feitosa, C.M., Melo, C.H.S., Santos, F.P.S., Silva, V.L., (2017). Atividades biológicas, farmacológicas e nutricionais de *Annona crassiflora* Mart. (Araticum)–estado da arte. Revista Brasileira de Plantas Medicinais, 19, 374–381.

Ferraz-Almeida, R., Da Mota, R.P., (2021). Routes of soil uses and conversions with the main crops in brazilian Cerrado: A scenario from 2000 to 2020. L. 10, 1135. https://doi.org/10.3390/land10111135

Fišer Pečnikar, Ž., Buzan, E.V., (2014). 20 years since the introduction of DNA barcoding: from theory to application. Journal of applied genetics, 55(1), 43-52. https://doi.org/10.1007/s13353-013-0180-y

Flora e Funga do Brasil. Jardim Botânico do Rio de Janeiro. Disponível em: < http://floradobrasil.jbrj.gov.br/ >. Acesso em: 13 Jan 2023

Ford, C.S., Ayres, K.L., Toomey, N., Haider, N., Stahl, J.V.A., Kelly, L.J., Wikström, N., Hollingsworth, P. M., Duff, R.J., Hoot, S.B., Cowan, R.S., Chase, M.W., Wilkinson, M.J., (2009). Selection of candidate coding DNA barcoding regions of use on land plants. Botanical Journal of the Linnean Society, 159, 1–11. https://doi.org/10.1111/j.1095-8339.2008.00938.x

Forzza, R.C., Baumgratz, J.F.A., Bicudo, C.E.M., Canhos, D.A.L, Carvalho Júnior, A.A., Costa, A., Costa, D.P., Hopkins, M., Leitman, P.M., Lohmann, L.G., Lughadha, E.N., Maia, L.C., Martinelli, G., Menezes, M., Morim, M.P., Nadruz Coelho, M.A., Peixoto, A.L., Pirani, J.R., Prado, J., Queiroz, L.P., Souza, S., Souza, V.C., Stehmann, J.R., Sylvestre, L.S., Walter, B.M.T, Zappi D., (2010). Preparation of a list of Brazilian plant and fungal species: methods and results. CABI Digital Library 1–4. https://www.cabidigitallibrary.org/doi/full/10.5555/20133083557 (accessed 7 June 2025).

Françoso, R.D., Dexter, K.G., Machado, R.B., Pennington, R.T., Pinto, J.R.R., Brandão, R.A., Ratter, J.A., (2020). Delimiting floristic biogeographic districts in the Cerrado and assessing their

conservation status. Biodiversity and Conservation, 29, 1477–1500. https://doi.org/10.1007/s10531-019-01819-3

Gibson, J.F., Shokralla, S., Curry, C., Baird, D.J., Monk, W.A., King, I., Hajibabaei, M., (2015). Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. PloS one, 10(10), e0138432.

Grant, D.M., Brodnicke, O.B., Evankow, A.M., Ferreira, A.O., Fontes, J.T., Hansen, A.K., Ekrem, T., (2021). The future of DNA barcoding: Reflections from early career researchers. Diversity, 13(7), 313. https://doi.org/10.3390/d13070313

Gostel, M.R., Kress, W.J., (2022). The expanding role of DNA barcodes: Indispensable tools for ecology, evolution, and conservation. Diversity, 14(3), 213. https://doi.org/10.3390/d14030213

González, A.V., Gómez-Silva, V., Ramírez, M.J., Fontúrbel, F.E., (2020). Meta-analysis of the differential effects of habitat fragmentation and degradation on plant genetic diversity. Conservation Biology, 34, 711–720. https://doi.org/10.1111/cobi.13422

Guerra, A., de Oliveira Roque, F., Garcia, L.C., Ochoa-Quintero, J.M., de Oliveira, P.T.S., Guariento, R. D., Rosa, I.M., (2020). Drivers and projections of vegetation loss in the Pantanal and surrounding ecosystems. Land use policy, 91, 104388. https://doi.org/10.1016/j.landusepol.2019.104388

Hajibabaei, M., Singer, G.A., Hebert, P.D., Hickey, D.A., (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. TRENDS in Genetics, 23, 167–172. https://doi.org/10.1016/j.tig.2007.02.001

Hamid, B.A., Al-Rufaye, Z.T.A., Khaleel, H.A., & Abidukadhim, A.R., (2023). Phylogenetic analysis of some species of cyperaceae based on DNA internal transcribed spacer (ITS) in Iraq. SABRAO J. Breed. Genet, 55(3), 708-716. http://doi.org/10.54910/sabrao2023.55.3.9

Hamilton, M., (1999). Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Molecular Ecology. 8, 521–523.

Hammer, Ø., Harper, D.A., (2001). Past: paleontological statistics software package for educaton and data anlysis. Palaeontologia electronica, 4(1), 1.

Hawkins, J., De Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., Adams-Groom, B., (2015). Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. PloS one, 10(8), e0134735. https://doi.org/10.1371/journal.pone.0134735

Hebert, P.D.N., Cywinska, A., Ball, S.L., Dewaard, J.R., (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society B, 270, 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert, P.D., Gregory, T.R., (2005). The promise of DNA barcoding for taxonomy. Systematic biology, 54, 852–859. https://doi.org/10.1080/10635150500354886

Hollingsworth, P.M., Graham, S.W., Little, D.P., (2011). Choosing and using a plant DNA barcode. PloS One, 6, e19254. https://doi.org/10.1371/journal.pone.0019254

Honer, M., (1993). Precipitação pluviométrica na região de Campo Grande, MS: uma análise do período 1970-1991. Campo Grande, MS: EMBRAPA-CNPGC. https://core.ac.uk/download/pdf/33883998.pdf (accessed 7 June 2025).

Hosein, F.N., Austin, N., Maharaj, S., Johnson, W., Rostant, L., Ramdass, A.C., Rampersad, S.N., (2017). Utility of DNA barcoding to identify rare endemic vascular plant species in Trinidad. Ecology and Evolution, *7*(18), 7311-7333. https://doi.org/10.1002/ece3.3220

Hoveka, L.N., Van der Bank, M., Boatwright, J.S., Bezeng, B.S., Yessoufou, K., (2016). The noncoding *trnH-psbA* spacer, as an effective DNA barcode for aquatic freshwater plants, reveals prohibited invasive species in aquarium trade in South Africa. South African Journal of Botany, 102, 208-216. https://doi.org/10.1016/j.sajb.2015.06.014

IMASUL Instituto de Meio Ambiente de Mato Grosso do Sul (2024). Available at: https://www.imasul.ms.gov.br/legislacao-ambiental/leis/ (last accessed: 15/05/2024).

Irigaray, C.T.J.H., da Silva, C.J., da Cunha, C.N., Calheiros, D.F., Girard, P., Junk, W.J., (2017). Contribuição técnico-científica ao aprimoramento do marco regulatório visando à proteção do Pantanal Mato-grossense. Pantanal, 127.

Jamdade, R., Upadhyay, M., Shaer, K.A., Harthi, E.A., Sallani, M.A., Jasmi, M.A., Ketbi, A.A., (2021). Evaluation of Arabian Vascular Plant Barcodes (*rbcL* and *matK*): Precision of Unsupervised and Supervised Learning Methods towards Accurate Identification. Plants, 10, 2741. https://doi.org/10.3390/plants10122741

Jeunen, G.J., Dowle, E., Edgecombe, J., von Ammon, U., Gemmell, N.J., Cross, H., (2023). crabs—A software program to generate curated reference databases for metabarcoding sequencing data. Molecular Ecology Resources, 23(3), 725-738. https://doi.org/10.1111/1755-0998.13741

Johnson, M.D., Fokar, M., Cox, R.D., Barnes, M.A., (2021). Airborne environmental DNA metabarcoding detects more diversity, with less sampling effort, than a traditional plant community survey. BMC Ecology and Evolution, 21(1), 218. https://doi.org/10.1186/s12862-021-01947-x

Joly, S., Davies, T.J., Archambault, A., Bruneau, A., Derry, A., Kembel, S.W., Peres-Neto, P., Vamosi, J., Wheeler, T.A., (2014). Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. Molecular Ecology Resources, 14, 221–232. https://doi.org/10.1111/1755-0998.12173

Junk, W.J., Da Cunha, C.N., Wantzen, K.M., Petermann, P., Strüssmann, C., Marques, M.I., Adis, J. (2006). Biodiversity and its conservation in the Pantanal of Mato Grosso, Brazil. Aquatic Sciences, 68(3), 278-309.

Kamenova S., Albon S.D., Loe L.E., Irvine R.J., Langvatn R., Gusarova G., de Muinck E.J., Trosvik P., (2025). Arctic Greening Drives Changes in the Diet and Gut Microbiome of a Large Herbivore With Consequences for Body Mass. Ecology and Evolution 15(7):e71731. https://doi.org/10.1002/ece3.71731

Kartzinel, T R., Hoff, HK., Divoll, T.J., Littleford-Colquhoun, B.L., Anderson, H.., Burak, M.K., Kuzmina, M.L., Musili, P.M., Rogers, H., Troncoso, A.J., Kartzinel R.Y.,. (2025). Global availability of plant DNA barcodes as genomic resources to support basic and policy-relevant biodiversity research. Molecular Ecology, 34(7), e17712.https://doi.org/10.1111/mec.17712

Keck, F., Couton, M., Altermatt, F., (2023). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Moloecular Ecology Resources, 23, 742–755. https://doi.org/10.1111/1755-0998.13746

Kim, K.J., Lee, H.L., (2005). Widespread occurrence of small inversions in the chloroplast genomes of land plants. Molecules and Cells, 19, 104–113. https://doi.org/10.1016/S1016-8478(23)13143-8

Klink, C.A., Sato, M.N., Cordeiro, G.G., Ramos, M.I.M., (2020). The role of vegetation on the dynamics of water and fire in the cerrado ecosystems: Implications for management and conservation. Plants, 9, 1803. https://doi.org/10.3390/plants9121803

Köppen, W. (1948). Climatología: Com un estudio de los climas de la tierra. Fondo de Cultura Econômica, México.

Kowalska, Z., Pniewski, F., Latala, A. (2019). DNA barcoding–A new device in phycologist's toolbox. Ecohydrology & Hydrobiology, 19, 417–427. https://doi.org/10.1016/j.ecohyd.2019.01.002

Kraaijeveld, K., De Weger, L.A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P.S., Den Dunnen, J.T., (2015). Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Molecular ecology resources, 15(1), 8-16. https://doi.org/10.1111/1755-0998.12288

Kress, W.J., Erickson, D.L., (2007). A two-locus global DNA barcode for land plants: the coding *rbcl* gene complements the non-coding *trnH-psbA* spacer region. PloS One, 2, e508. https://doi.org/10.1371/journal.pone.0000508

Kress, W.J., Erickson, D.L., Swenson, N.G., Thompson, J., Uriarte, M., Zimmerman, J.K., (2010). Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot, PloS One 5, e15409. https://doi.org/10.1371/journal.pone.0015409

Kress, W.J., García-Robledo, C., Uriarte, M., Erickson, D. L., (2015). DNA barcodes for ecology, evolution, and conservation. Trends in ecology & evolution, 30(1), 25-35. http://dx.doi.org/10.1016/j.tree.2014.10.008

Kress, W.J., (2017). Plant DNA barcodes: Applications today and in the future. Journal of systematics and evolution, 55, 291–307. https://doi.org/10.1111/jse.12254

Leal Filho, W., Azeiteiro, U.M., Salvia, A.L., Fritzen, B., Libonati, R., (2021). Fire in Paradise: Why the Pantanal is burning. Environmental Science & Policy, 123, 31-34. https://doi.org/10.1016/j.envsci.2021.05.005

Lee, S.Y., Ng, W.L., Mahat, M.N., Nazre, M., Mohamed, R., (2016). DNA barcoding of the endangered Aquilaria (Thymelaeaceae) and its application in species authentication of agarwood products traded in the market. PloS one, 11(4), e0154631. https://doi.org/10.1371/journal.pone.0154631

Letsiou, S., Madesis, P., Vasdekis, E., Montemurro, C., Grigoriou, M.E., Skavdis, G., Moussis, V., Koutelidakis, A.E., Tzakos, A.G., (2024). DNA Barcoding as a Plant Identification Method. Applied Sciences, 14, 1415. https://doi.org/10.3390/app14041415

Li, X., Yang, Y., Henry, R.J., Rosseto, M., Wang, Y., Chen, S., (2015). Plant DNA barcoding: from gene to genome. Bioogical Reviews, 90, 157–166. https://doi.org/10.1111/brv.12104

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. Bioinformatics, 37(23), 4572-4574. https://doi.org/10.1093/bioinformatics/btab705

Lima, D.C., da Rocha Alves, M., Noguera, N.H., do Nascimento, R.D., (2022). A review on Brazilian baru plant (*Dipteryx alata* Vogel): morphology, chemical composition, health effects, and technological potential. Future Foods 5, 100146. https://doi.org/10.1016/j.fufo.2022.100146

Lima, J.S., Collevatti, R.G., de Oliveira, L.K., Chaves, L.J., Naves, R.V., Soares, T.N., Filho, J.A.F.D., de Campos Telles, M.P., (2023). Forecasting effects of climate changes on the population genetic structure of Anacardium occidentale in the Cerrado biome, Brazil. Genetica, 151, 357–367. https://doi.org/10.1007/s10709-023-00199-5

Little, D.P., (2011). DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. PLoS one, 6(8), e20552. https://doi.org/10.1371/journal.pone.0020552

Liu, J.I.E., Moeller, M., Gao, L.M., Zhang, D.Q., Li, D.Z., (2011). DNA barcoding for the discrimination of Eurasian yews (Taxus L., Taxaceae) and the discovery of cryptic species. Molecular ecology resources, 11(1), 89-100. https://doi.org/10.1111/j.1755-0998.2010.02907.x

Liu, J., Shi, L., Han, J., Li, G., Lu, H., Hou, J., ... & Downie, S. R. (2014). Identification of species in the angiosperm family Apiaceae using DNA barcodes. Molecular ecology resources, 14(6), 1231-1238. https://doi.org/10.1111/1755-0998.12262

Liu, Z.F., Ci, X.Q., Li, L., Li, H.W., Conran, J.G., Li, J., (2017). DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China. PloS one, 12(4), e0175788. https://doi.org/10.1371/journal.pone.0175788

Liu, Y., Xu, C., Dong, W., Yang, X., Zhou, S., (2021). Determination of a criminal suspect using environmental plant DNA metabarcoding technology. Forensic Science International, 324, 110828. https://doi.org/10.1016/j.forsciint.2021.110828

Loera-Sánchez, M., Studer, B., Kölliker, R., (2020). DNA barcode *trnH-psbA* is a promising candidate for efficient identification of forage legumes and grasses. BMC research notes, 13(1), 35. https://doi.org/10.1186/s13104-020-4897-5

Maia, V. H., Mata, C. S. D., Franco, L. O., Cardoso, M. A., Cardoso, S. R. S., Hemerly, A. S., & Ferreira, P. C. G. (2012). DNA barcoding Bromeliaceae: achievements and pitfalls. PLoS one, 7(1), e29877. https://doi.org/10.1371/journal.pone.0029877

Marengo, J.A., Oliveira, G.S., Alves, L.M. (2015). Climate Change Scenarios in the Pantanal. In: Bergier, I., Assine, M. (eds) Dynamics of the Pantanal Wetland in South America. The Handbook of Environmental Chemistry, vol 37. Springer, Cham. https://doi.org/10.1007/698_2015_357

Martins, P.I., Belém, L.B.C., Peluso, L., Szabo, JK., Trindade, W.F., Pott, A., Garcia, L.C., (2024). Priority Areas for Integrated Fire Management and Ecological Restoration in the Upper Paraguay River Basin (Brazil) Based on Distributions of Fire-Sensitive and Threatened Plant Species. Available at SSRN 4736255. http://dx.doi.org/10.2139/ssrn.4736255

Maya-Lastra CA. Datataxa vU, 2019. https://github.com/camayal/Datataxa (accessed 26 May 2025).

Maya-Lastra, C.A., Alvarado-Cárdenas, L.O., Rodríguez-Gómez, F.C., Urrea-Galeano, L.A., Villaseñor J.L., Ruiz-Sanchez, E., (2022). The Mexican flora as a case study in systematics: a meta-analysis of GenBank accessions. Botanical Sciences, 100, 66–82. https://doi.org/10.17129/botsci.3061

Meier, R., Shiyang, K., Vaidya, G., Ng, P.K., (2006). DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic biology, 55(5), 715-728. https://doi.org/10.1080/10635150600969864

Michopoulos, P. (2011). Biogeochemistry of Urban Forests. In: Levia, D., Carlyle-Moses, D., Tanaka, T. (Eds.), Forest Hydrology and Biogeochemistry. Ecological Studies 216. Springer, Dordrecht, pp. 341–353. https://doi.org/10.1007/978-94-007-1363-5_16

Mir, R.A., Bhat, K.A., Rashid, G., Ebinezer, L.B., Masi, A., Rakwal, R., Shah, A.A., Zargar, M.S., (2021). DNA barcoding: a way forward to obtain deep insights about the realistic diversity of living organisms. The Nucleus, 64, 157–165. https://doi.org/10.1007/s13237-020-00330-3

Mitsch, W.J., Gosselink, J.G., Zhang, L., Anderson, C.J., (2009). Wetland ecosystems. John Wiley & Sons.

Mort, M.E., Crawford, D.J., Archibald, J.K., O'Leary, T.R., Santos-Guerra, A., (2010). Plant DNA barcoding: A test using Macaronesian taxa of *Tolpis* (Asteraceae). Taxon, 59, 581–587. https://doi.org/10.1002/tax.592022

Moretti, E.C., Moretti, S.A.L., (2018). Pantanal brasileiro, território vulnerável para quem? conflitos e desafios na produção de geografias. América Latina en las últimas décadas: procesos y retos, 17, 161.

Murray, D. C., Pearson, S.G., Fullagar, R., Chase, B.M., Houston, J., Atchison, J., Bunce, M., (2012). High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. Quaternary Science Reviews, 58, 135-145. https://doi.org/10.1016/j.quascirev.2012.10.021

Nath, S., Vanslambrouck, J.T., Yao, J.W., Gullapalli, A., Razi, F., Lu, Y., (2024). DNA barcoding of terrestrial invasive plant species in Southwest Michigan. Plant Direct, 8, e615. https://doi.org/10.1002/pld3.615

Oliveira, U., Paglia, A. P., Brescovit, A. D., de Carvalho, C. J., Silva, D. P., Rezende, D. T., ... & Santos, A. J. (2016). The strong influence of collection bias on biodiversity knowledge shortfalls of B razilian terrestrial biodiversity. Diversity and Distributions, 22(12), 1232-1244. https://doi.org/10.1111/ddi.12489

Orioli, L., Scariot, A., (2021). Land management for cattle farming affects the persistence of an important fruit tree in the Brazilian Savanna. Flora, 285, 151950. https://doi.org/10.1016/j.flora.2021.151950

Pawlowski, J., Apothéloz-Perret-Gentil, L., Altermatt, F., (2020). Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. Molecular ecology, 29(22), 4258-4264. https://doi.org/10.1111/mec.15643

Papadopoulou, A., Chesters, D., Coronado, I., de la Cadena, G., Cardoso, A., Reyes, J.C., Maes, J.M., Rueda, R.M., Gómez-Zurita, J., (2015). Automated DNA-based plant identification for large-scale biodiversity assessment. Molecular Ecology Resources, 15, 136–152. https://doi.org/10.1111/1755-0998.12256

PCBAP, B., (1997). Ministério do Meio Ambiente, dos Recursos Hídricos e da Amazônia Legal. Plano de Conservação da Bacia do Alto Paraguai (Pantanal)–PCBAP: Análise integrada e prognóstica da Bacia do Alto Paraguai. Programa Nacional do Meio Ambiente. Projeto Pantanal, 3, 369.

Pereira-Silva, L., Trevisan, R., Escudero, M., Alcantara, S., de Oliveira Ribeiro, A. R., Larridon, I., (2025). Biogeography and diversification dynamics of the megadiverse plant genus *Cyperus* (Cyperaceae, Poales). Annals of Botany, mcaf117. https://doi.org/10.1093/aob/mcaf117

Pires, A.C., Marinoni, L., (2010). DNA barcoding and traditional taxonomy unified through Integrative Taxonomy: a view that challenges the debate questioning both methodologies. Biota Neotropica, 10, 339–346. https://doi.org/10.1590/S1676-06032010000200035

Pott, A., Pott, V.J., (1994). Plantas do Pantanal (Plants of the Pantanal). Brasília: Embrapa.

Pott, A., Pott, V.J., (2022). Flora of the Pantanal. In Flora and vegetation of the Pantanal wetland (pp. 39-228). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-83375-6_3

Ratnasingham, S., Wei, C., Chan, D., Agda, J., Agda, J., Ballesteros-Mejia, L., Hebert, P.D., (2024). BOLD v4: A centralized bioinformatics platform for DNA-based biodiversity data. In DNA

barcoding: Methods and protocols (pp. 403-441). New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-3581-0_26

Ratter, J. A., Bridgewater, S., & Ribeiro, J. F. (2003). Analysis of the floristic composition of the Brazilian cerrado vegetation III: comparison of the woody vegetation of 376 areas. Edinburgh journal of botany, 60(1), 57-109. https://doi.org/10.1017/S0960428603000064

Reid, C.S., Doyle, V.P., Carter, J.R., Vargas-Rodriguez, Y., Urbatsch, L. E., (2017). Molecular systematics of targeted flat sedges (*Cyperus*, Cyperaceae) of the Americas. Plant Ecology and Evolution, 150(3), 343-357. https://doi.org/10.5091/plecevo.2017.1262

Rosa, C., Baccaro, F., Cronemberger, C., Hipolito, J., Barros, C., Rodrigues, D.D.J., Magnusson, W.E., (2021). The Program for Biodiversity Research in Brazil: The role of regional networks for biodiversity knowledge, dissemination, and conservation. Anais da Academia Brasileira de Ciências, 93, e20201604. https://doi.org/10.1590/0001-3765202120201604

Rubinoff, D., Cameron, S., Will, K., (2006). Are plant DNA barcodes a search for the Holy Grail? Trends in Ecology and Evolution, 21, 1–2. https://doi.org/10.1016/j.tree.2005.10.019

Rojas, V.M., Marconi, L.F.C.B., Guimaraes-Inacio, A., Leimann, F.V., Tanamati, A., Gozzo, A. M., Fuchs, R.H.B., Barreiro, M.F., Barros, L., et al., (2019). Formulation of mayonnaises containing PUFAs by the addition of microencapsulated chia seeds, pumpkin seeds and baru oils. Food Chemistry, 274, 220–227. https://doi.org/10.1016/j.foodchem.2018.09.015

Roy, A., Frascaria, N., Mackay, J., Bousquet, J., (1992). Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis.* Theoretical and Applied Genetics, 85(2), 173-180. https://doi.org/10.1007/BF00222856

Sahu, A., Kumar, N., Singh, C.P., Singh, M., (2023). Environmental DNA (eDNA): Powerful technique for biodiversity conservation. Journal for Nature Conservation, 71, 126325. https://doi.org/10.1016/j.jnc.2022.126325

Salgotra, R.K., Chauhan, B.S. (2023). Genetic diversity, conservation, and utilization of plant genetic resources. Genes, 14(1), 174. https://doi.org/10.3390/genes14010174

Sang, T., Crawford, D.J., Stuessy, T.F., (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). The American journal of Botany, 84, 1120–1136. https://doi.org/10.2307/2446155

Sartori, Â.L.B., Pott, V.J., Pott, A., de Carvalho, F.S., (2018). Check-list das Angiospermas do Chaco de Mato Grosso do Sul. Iheringia, Série Botânica., 73, 22-33. https://doi.org/10.21826/2446-8231201873s22

Schmidt, I.B., Moura, L.C., Ferreira, M.C., Eloy, L., Sampaio, A.B., Dias, P.A., Berlinck, C.N., (2018). Fire management in the Brazilian savanna: First steps and the way forward. Journal of applied ecology, 55(5), 2094-2101. https://doi.org/10.1111/1365-2664.13118

Schoch, C.L., Ciufo, S., Domrachev, M., Hotton, C.L., Kannan, S., Khovanskaya, R., Karsch-Mizrachi, I., (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database, 2020, baaa062. https://doi.org/10.1093/database/baaa062

Seidl, A. F., & Moraes, A. S. (2000). Global valuation of ecosystem services: application to the Pantanal da Nhecolandia, Brazil. Ecological economics, 33(1), 1-6. https://doi.org/10.1016/S0921-8009(99)00146-9

Setsuko, S., Yoshimura, K., Ueno, S., Worth, J.R.P., Ujino-Ihara, T., Katsuki, T., Yoshimaru, H., (2023). A DNA barcode reference library for the native woody seed plants of Japan. Molecular Ecology Resources, 23(4), 855-871. https://doi.org/10.1111/1755-0998.13748

Shehzad, W., Riaz, T., Nawaz, M. A., Miquel, C., Poillot, C., Shah, S. A., ... & Taberlet, P. (2012). Carnivore diet analysis based on next-generation sequencing: Application to the leopard cat (Prionailurus bengalensis) in Pakistan. Molecular ecology, 21(8), 1951-1965. https://doi.org/10.1111/j.1365-294X.2011.05424.x

Silva, J.S.V., Pott, A., Abdon, M.M., Pott, V.J. Santos, K.R., (2011). Projeto GeoMS: cobertura vegetal e uso da terra do Estado de Mato Grosso do Sul. Empresa Brasileira de Pesquisa Agropecuária. Informática Agropecuária, Campinas.

Silveira, P., dos Santos, J.S., Pena, J.C., de Lima, N.E., Vitorino, L.C., Martello, F., Guimaraes, R.A., Moreira, J.W., Gomes, J.C., Araújo, L.M., Ribeiro, M.C., Collevatti, R.G., (2025). The loss of an unknown biodiversity: Spatial gaps in plant survey and conservation in a Brazilian hotspot of biodiversity. Biological Conservation, 305, 111098. https://doi.org/10.1016/j.biocon.2025.111098

Sobral, M., Stehmann, J.R., (2009). An analysis of new angiosperm species discoveries in Brazil (1990-2006). Taxon, 58, 227–232. https://doi.org/10.1002/tax.581021

Song, J., Shi, L., Li, D., Sun, Y., Niu, Y., Chen, Z., Luo, H., Pang, X., Sun, Z., et al., (2012). Extensive pyrosequencing reveals frequent intra-genomic variations of internal transcribed spacer regions of nuclear ribosomal DNA. PloS One, 7, 43971 https://doi.org/10.1371/journal.pone.0043971

Souza Jr, C.M., Z. Shimbo, J., Rosa, M.R., Parente, L.L., A. Alencar, A., Rudorff, B.F., Azevedo, T., (2020). Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. Remote Sensing, 12(17), 2735. https://doi.org/10.3390/rs12172735

Srirama, R., Senthilkumar, U., Sreejayan, N., Ravikanth, G., Gurumurthy, B.R., Shivanna, M.B., Shaanker, R.U., (2010). Assessing species admixtures in raw drug trade of Phyllanthus, a hepato-protective plant using molecular tools. Journal of ethnopharmacology, 130(2), 208-215. https://doi.org/10.1016/j.jep.2010.04.042

Stepien, C. A., Lance, R. F., Klymus, K. E., & Hunter, M. E. (2023). The Government eDNA Working Group 6th Annual eDNA Technical Exchange Workshop. Environmental DNA, 5(6), 1196-1201. https://doi.org/10.1002/edn3.466

Strassburg, B.B., Iribarrem, A., Beyer, H.L., Cordeiro, C.L., Crouzeilles, R., Jakovac, C.C., Junqueira, A.B., Lacerda, E., Latawiec, A.E., et al (2020). Global priority areas for ecosystem restoration. Nature, 586, 724–729. https://doi.org/10.1038/s41586-020-2784-9

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., ... & Willerslev, E. (2007). Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. Nucleic acids research, *35*(3), e14-e14. https://doi.org/10.1093/nar/gkl938

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. Molecular ecology, 21(8). https://doi.org/10.1111/j.1365-294X.2012.05542.x

Tahir, A., Hussain, F., Ahmed, N., Ghorbani, A., Jamil, A., (2018). Assessing universality of DNA barcoding in geographically isolated selected desert medicinal species of Fabaceae and Poaceae. PeerJ, 6, e4499. https://doi.org/10.7717/peerj.4499

Tan, S.L., Luo, Y.H., Hollingsworth, P.M., Burgess, K.S., Xu, K., Li, D Z., Gao, L.M., (2018). DNA barcoding herbaceous and woody plant species at a subalpine forest dynamics plot in Southwest China. Ecology and Evolution, 8(14), 7195-7205. https://doi.org/10.1002/ece3.4254

Thomas, W.W., Magill, R.E., (2002). Uma proposta para um check-list on line das plantas neotropicais. In: Araújo, E.L., Moura, N.A., Sampaio, E.V.S.B., Gestinari, L.M.S., Carneiro, J.M.T. (Eds.),

Biodiversidade, conservação e uso sustentável da flora do Brasil. LIII Congresso Nacional de Botânica/XXV Reunião Nordestina de Botânica. Recife, pp. 252–254.

Tnah, L.H., Lee, S.L., Lee, C.T., Ng, K.K.S., Ng, C.H., Zawiah, N., (2024). DNA barcode identification of cultivated and wild tropical fruit species. 3 Biotech, 14(1), 7. https://doi.org/10.1007/s13205-023-03848-w

Trinugroho, J.P., Sudigyo, D., Hidayat, A.A., Nirwantono, R., Mahesworo, B., Pardamean, B., (2023, May). Identification of rice plants via DNA barcoding for securing future food availability. In IOP Conference Series: Earth and Environmental Science (Vol. 1183, No. 1, p. 012033). IOP Publishing.

Trujillo-Argueta, S., Del Castillo, R.F., Tejero-Diez, D., Matias-Cervantes, C.A., Velasco-Murguía, A., (2021). DNA barcoding ferns in an unexplored tropical montane cloud forest area of southeast Oaxaca, Mexico. Scientific Reports, 11(1), 22837. https://doi.org/10.1038/s41598-021-02237-8

Trujillo-Argueta, S., Del Castillo, R.F., Velasco-Murguía, A., (2022). Testing the effectiveness of rbcLa DNA-barcoding for species discrimination in tropical montane cloud forest vascular plants (Oaxaca, Mexico) using BLAST, genetic distance, and tree-based methods. PeerJ, 10, e13771. https://doi.org/10.7717/peerj.13771

Torti, S. D., Coley, P. D., & Kursar, T. A. (2001). Causes and consequences of monodominance in tropical lowland forests. The american naturalist, 157(2), 141-153. https://doi.org/10.1086/318629

Valentini, A., Pompanon, F., Taberlet, P., (2009). DNA barcoding for ecologists. Trends in ecology & evolution, 24(2), 110-117.

Vasar, M., Davison, J., Moora, M., Sepp, S.K., Anslan, S., Al-Quraishy, S., Zobel, M., (2023). Metabarcoding of soil environmental DNA to estimate plant diversity globally. Frontiers in plant science, *14*, 1106617. https://doi.org/10.3389/fpls.2023.1106617

Vasconcelos, S., Nunes, G.L., Dias, M.C., Lorena, J., Oliveira, R.R.M., Lima, T.G.L., Pires, E.S., Valadares, R.B.S., Alves, R., Watanabe, M.T.C., Zappi, D.C., Hiura, A.L., Pastore, M., Vasconcelos, L.V., Mota, N.F.O., Viana, P.L., Gil, A.S.B., Simões, A.O., Imperatriz-Fonseca, V.L., Harley, R.M., Giulietti, A.M., Oliveira, G., (2021). Unraveling the plant diversity of the Amazonian canga through DNA barcoding. Ecology and Evolution, 11, 13348–13362. https://doi.org/10.1002/ece3.8057

Xu, S.Z., Li, Z.Y., Jin, X.H., (2018). DNA barcoding of invasive plants in China: A resource for identifying invasive plants. Molecular ecology resources, 18(1), 128-136. https://doi.org/10.1111/1755-0998.12715

Wang, Y., Pedersen, M.W., Alsos, I.G., De Sanctis, B., Racimo, F., Prohaska, A., Willerslev, E., (2021). Late Quaternary dynamics of Arctic biota from ancient environmental genomics. Nature, 600(7887), 86-92. https://doi.org/10.1038/s41586-021-04016-x

Wang, J., Wang, X., Liu, L., Wang, X., Wang, J., Zheng, Y., Pan, X., (2024). Analyzing the Interaction between Tetrahymena pyriformis and Bacteria under Different Physicochemical Conditions When Infecting Guppy Using the eDNA Method. Animals: an Open Access Journal from MDPI, 14(15), 2194. https://doi.org/10.3390/ani14152194

Ward, Tockner, (2001). Biodiversity: towards a unifying theme for river ecology. Freshwater Biology, 46(6), 807-819. https://doi.org/10.1046/j.1365-2427.2001.00713.x

Wardani, I.G.A.K., Amandita, F.Y., de Melo Moura, C.C., Gailing, O., Siregar, I.Z., (2022). Molecular taxonomy via DNA barcodes for species identification in selected genera of Fabaceae. Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (Journal of Natural Resources and Environmental Management), 12(1), 112-122. https://doi.org/10.29244/jpsl.12.1.112-122

Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., ... & Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Science of the Total Environment, 678, 499-524. https://doi.org/10.1016/j.scitotenv.2019.04.247

West, K.M., Stat, M., Harvey, E.S., Skepper, C.L., DiBattista, J.D., Richards, Z.T., Travers, M.J., Newman, S.J., Bunce, M. (2020). eDNA metabarcoding survey reveals fine-scale coral reef community variation across a remote, tropical island ecosystem. Molecular Ecology, 29, 1069–1086. https://doi.org/10.1111/mec.15382

Winter, M., Devictor, V., Schweiger, O., (2013). Phylogenetic diversity and nature conservation: where are we? Trends in Ecology and Evolution, 28, 199–204. http://dx.doi.org/10.1016/j.tree.2012.10.015

Yang, C.Q., Lv, Q., Zhang, A.B., (2020). Sixteen years of DNA barcoding in China: What has been done? What can be done?. Frontiers in Ecology and Evolution, 8, 57. https://doi.org/10.3389/fevo.2020.00057

Yang, J., Yang, J., Xing, D., Luo, X., Lu, S., Huang, C., Hahs, A.K., (2021). Impacts of the remnant sizes, forest types, and landscape patterns of surrounding areas on woody plant diversity of urban remnant forest patches. Urban Ecosystems, 24, 345–354. https://doi.org/10.1007/s11252-020-01040-z

Yoccoz, N.G., Bråthen, K.A., Gielly, L., Haile, J., Edwards, M.E., Goslar, T., Taberlet, P., (2012). DNA from soil mirrors plant taxonomic and growth form diversity. Molecular ecology, 21(15), 3647-3655. https://doi.org/10.1111/j.1365-294X.2012.05545.x