UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Pós-Graduação em Zoologia

Bárbara Regina Neves Chaves

Metabarcodes de iDNA para Caracterização da Diversidade de Mamíferos no Brasil

Belo Horizonte
2024

Bárbara Regina Neves Chaves

METABARCODES DE IDNA PARA CARACTERIZAÇÃO DA DIVERSIDADE DE MAMÍFEROS NO BRASIL

Tese apresentada ao programa de Pós-Graduação em Zoologia da Universidade Federal de Minas Gerais como requisito parcial para obtenção do grau de Doutora em Zoologia.

Orientador: Prof. Dr. Fabrício Rodrigues dos Santos

Belo Horizonte
2024

043 Chaves, Bárbara Regina Neves.
Metabarcodes de iDNA para caracterização da diversidade de mamíferos no Brasil [manuscrito] / Bárbara Regina Neves Chaves. – 2024.
50 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Fabrício Rodrigues dos Santos.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Zoologia.

1. Zoologia. 2. Mamíferos. 3. Ecossistema. 4. Código de Barras de DNA Taxonômico. I. Santos, Fabrício Rodrigues dos. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 591

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis – CRB: 6/2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Metabarcodes de iDNA para caracterização da diversidade de mamíferos no Brasil

BÁRBARA REGINA NEVES CHAVES

Esta tese foi apresentada em sessão pública e submetida a avaliação em 19 de dezembro de 2024, sendo aprovada pela Banca Examinadora composta pelos seguintes membros:

Prof. Dr. Bruno Henrique Saranholi (Membro / UFSCAR)

Prof. Dr. Fabrício Rodrigues dos Santos (Orientador / UFMG)

Profa. Dra. Izabela Santos Mendes (Membro / ITV)

Profa. Dra. Karine Frehner Kavalco (Membro / UFV)

Prof. Dr. Pedro Manoel Galetti Júnior (Membro / UFSCAR)

Documento assinado eletronicamente por **Bruno Henrique Saranholi**, **Usuário Externo**, em 03/01/2025, às 16:42, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Fabricio Rodrigues dos Santos**, **Professor do Magistério Superior**, em 03/01/2025, às 17:08, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Pedro Maanoel Galetti Junior**, **Usuário Externo**, em 04/01/2025, às 10:23, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Izabela Santos Mendes**, **Usuário Externo**, em 06/01/2025, às 09:57, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Karine Frehner Kavalco**, **Usuário Externo**, em 06/01/2025, às 19:31, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Agradecimentos

Ao Fabrício, que manteve a porta aberta para o meu retorno, confiando no meu trabalho e me propondo esse desafio. À Patrícia, ao Greg e ao Centro de Microscopia, que gentilmente proporcionaram minha dedicação exclusiva a este trabalho nos últimos dois anos e meio. Ao José Eustáquio, que colaborou com a ideia e a realização de todas as coletas de campo. Aos colegas do LBEM, que ao longo desses quatro anos e meio, ou mesmo antes, ajudaram nas coletas, na bancada e nas telas pretas, em especial Pedro, Mateus, Julia, Ana Cristina, Jean, Davidson, Henry, Joana e Maria Eugênia (*in memorian*). À Lica e ao Fred, que cederam espaço no Laboratório de Sistemática de Insetos e me ajudaram na triagem das amostras. Ao Rennan, que me acudiu e fez o sequenciamento das amostras no Laboratório de Genômica, e ao Renato Oliveira, que pacientemente me ajudou nas análises. Ao Tulaci, que ajudou principalmente nos momentos finais de desespero e escrita, e desenhando os mapas deste trabalho. À CAPES, por dois anos de bolsa, e ao CNPq e à FAPEMIG, pelo financiamento do trabalho. Finalmente, à UFMG, minha *alma mater* e meu local de trabalho, pela estrutura física e suporte financeiro.

Resumo

A sistemática e a taxonomia são essenciais para a classificação biológica, mas enfrentam limitações no uso de caracteres morfológicos, especialmente em ecossistemas neotropicais, que abrigam uma rica biodiversidade, são altamente ameaçados e subamostrados, resultando na extinção de muitas espécies antes de serem descritas. Técnicas moleculares, como o *metabarcoding*, oferecem identificações rápidas e precisas, sem a necessidade de conhecimento prévio de morfologia. O *metabarcoding* identifica espécies a partir de fragmentos curtos de DNA liberados pelos organismos no ambiente. Sua eficiência é similar ou superior à das câmeras-trap, que, embora úteis em levantamentos de mamíferos, falham em identificar espécies pequenas ou arborícolas. Dípteros hematófagos e saprófagos possuem alta capacidade de dispersão e podem localizar mamíferos furtivos em ambientes tropicais, com sua coleta sendo mais eficiente por meio de armadilhas de interceptação de voo. A abordagem de *metabarcoding* de iDNA permite identificar moléculas de DNA de mamíferos a partir do conteúdo intestinal desses insetos. Como o *metabarcoding* não possui um marcador genético padrão, é recomendado o uso de múltiplos marcadores para melhorar a resolução taxonômica. Esta tese testou a eficácia do *metabarcoding* de iDNA em três biomas brasileiros (Amazônia, Mata Atlântica e Cerrado), adaptando primers à diversidade neotropical para melhorar a detecção de espécies e utilizando moscas das famílias Calliphoridae, Sarcophagidae e Tabanidae. Esta tese destaca a importância de integrar o *metabarcoding* às estratégias de conservação e biomonitoramento, otimizando a análise da biodiversidade e apoiando políticas públicas para a preservação dos biomas brasileiros.

Palavras-chave: *metabarcoding;* idna; biomas neotropicais; conservação de mamíferos.

Abstract

Systematics and taxonomy are essential for biological classification, but they face limitations when using only morphological characters, especially in Neotropical ecosystems, which host rich biodiversity, are highly threatened, and are under-sampled, leading to the extinction of many species before they are described. Molecular techniques, such as metabarcoding, provide rapid and accurate identifications without the need for prior morphological knowledge. Metabarcoding identifies species from short DNA fragments released by organisms into the environment. Its efficiency is similar to or greater than that of camera traps, which, although useful in mammal surveys, fail to identify small or arboreal species. Hematophagous and saprophagous dipterans have a high dispersal capacity and can locate elusive mammals in tropical environments, with their collection being more efficient through flight interception traps. The iDNA metabarcoding approach allows the identification of mammal DNA molecules from the intestinal contents of these insects. Since metabarcoding does not have a standard genetic marker, the use of multiple markers is recommended to improve taxonomic resolution. This thesis tested the effectiveness of iDNA metabarcoding in three Brazilian biomes (Amazon, Atlantic Forest, and Cerrado), adapting primers to Neotropical diversity to improve species detection, using flies from the families Calliphoridae, Sarcophagidae, and Tabanidae. This thesis highlights the importance of integrating metabarcoding into conservation and biomonitoring strategies, optimizing biodiversity analysis and supporting public policies for the preservation of Brazilian biomes.

Keywords: metabarcoding; idna; neotropical biomes; mammal conservation.

Sumário

1. Introdução

O número de espécies formalmente descritas é estimado entre 1,5 e 2 milhões (Larsen et al., 2017), mas calcula-se que cerca de 86% das espécies ainda sejam desconhecidas, especialmente em regiões tropicais (Dirzo & Raven, 2003). O número real de espécies pode ser muito maior, chegando a cerca de 8,7 milhões (Mora et al., 2011). No entanto, a maioria dos indicadores sobre o estado da natureza está em declínio, incluindo o número de espécies e o tamanho de suas populações, com projeções indicando que essa situação pode piorar nas próximas décadas, a menos que ações rápidas e integradas sejam implementadas (Diaz et al., 2019). O problema é ainda mais acentuado em regiões reconhecidas como *hotspots* de biodiversidade, uma vez que abrigam excepcional riqueza de espécies, ao mesmo tempo que enfrentam graves ameaças das atividades humanas (Myers et al 2000; Jenkins et al 2013). É o caso de biomas neotropicais, considerados entre os mais degradados e em rápido desaparecimento do mundo. Apesar de sua importância, esses ecossistemas são desproporcionalmente subamostrados (Hughes et al 2021), e inúmeras espécies neotropicais sofrem extinções locais ou desaparecem antes de serem formalmente descritas. Isso ressalta a importância de se ampliar o conhecimento taxonômico para compreender e conservar a biodiversidade.

A sistemática tem sido o alicerce da classificação biológica por séculos, fornecendo um sistema de referência para toda a biologia. Os principais objetivos da sistemática são descrever, classificar, nomear e determinar as relações da biodiversidade da Terra (Teletchea, 2016), além de fornecer levantamentos e inventários por meio de sua subdisciplina chamada taxonomia (Wilson, 2004; Crisci, 2006). Tradicionalmente, a sistemática e a taxonomia baseiam-se no exame de caracteres observáveis dos organismos, como estrutura corporal, coloração e características anatômicas. Muitas vezes são necessários equipamentos e ensaios específicos para detecção de certos caracteres morfológicos, como microscopia óptica e eletrônica ou mesmo tomografia computadorizada (Tessler et al 2022). No entanto, a sistemática tradicional frequentemente enfrenta limitações relacionadas à disponibilidade de caracteres morfológicos, que podem variar devido a fatores ambientais, estágios de desenvolvimento ou convergência evolutiva (Teletchea, 2016).

Diante das complexidades da diversidade biológica, que não se refletem apenas na morfologia, técnicas modernas surgiram não para substituir os estudos morfológicos, mas para oferecer uma compreensão mais abrangente da diversidade biológica (Hillis, 1987). Nesse contexto, técnicas moleculares têm se mostrado ferramentas poderosas para a classificação taxonômica, utilizando sequências de DNA como fonte de caracteres em análises filogenéticas, permitindo identificar homologias, apomorfias e plesiomorfias entre organismos (Wägele, 1995). Essas técnicas também revelam informações frequentemente ignoradas por métodos tradicionais, como diversidades intraespecíficas e espécies crípticas. Baseados no uso desses caracteres moleculares, o DNA barcoding foi proposto em 2003 como uma alternativa para superar limitações da classificação tradicional baseada em morfologia, permitindo uma identificação rápida e precisa de espécies, especialmente para não taxonomistas (Hebert et al., 2003).

O DNA barcoding se mostra uma ferramenta valiosa para a identificação de espécies em alta escala sem a necessidade de conhecimento prévio sobre morfologia, tendo transformado de forma significativa os inventários de biodiversidade nas últimas décadas (Miller et al., 2016). A identificação das espécies é realizada através da análise da diversidade de sequências de nucleotídeos de uma região única do DNA (o marcador genético) que apresente variação rápida o suficiente para diferenciar linhagens com isolamento reprodutivo recente (Felsenstein 2004). Uma região de 648 pares de bases do gene mitocondrial Citocromo Oxidase Subunidade I (COI) foi definida como o marcador genético padrão para DNA barcoding de metazoários (Hebert et al 2003) e tem sido descrita para uma ampla diversidade de organismos.

O DNA pode ser efetivamente extraído a partir de tecidos obtidos diretamente dos espécimes de interesse (p.ex. Mota el al. 2018; Pereira et al. 2011; Ilunga et al. 2020), mas essa não é a única forma

de amostragem de material genético. Como organismos vivos continuamente liberam moléculas de DNA no ambiente, estas podem ser utilizadas como indícios de sua presença em um local específico. Mais recentemente, a abordagem de *metabarcoding* surgiu como uma evolução do DNA barcoding, permitindo a identificação de espécies a partir do chamado DNA ambiental (eDNA), que consiste em uma mistura complexa de material genético liberado pelos organismos vivos no ambiente (Haile et al., 2009; Taberlet et al., 2012; Andújar et al., 2018). Devido à natureza mais fragmentada do eDNA, o *metabarcoding* geralmente amplifica e sequencia fragmentos de DNA ainda mais curtos utilizando técnicas de sequenciamento de alto rendimento (HTS) (Taberlet et al., 2012).

Ao contrário do DNA barcoding, não há um marcador genético padrão para o *metabarcoding* de vertebrados. Embora a região do COI utilizada no DNA barcoding pudesse ser adequada, ela é considerada muito longa para o fragmentado eDNA, e sua alta taxa evolutiva resulta na falta de sítios conservados para primers (Kent, 2009). Na ausência de um marcador genético padrão, diferentes marcadores e pares de primers podem ser escolhidos para estudos de *metabarcoding*. Por exemplo, pequenos fragmentos de genes ribossômicos, como RNA12S e RNA16S, são frequentemente utilizados devido aos seus sítios de primer altamente conservados, que permitem a amplificação de uma ampla variedade de táxons (Green et al., 2015). Para aumentar o poder de resolução, já que sequências muito curtas podem não apresentar nucleotídeos espécie-específicos (Freeland, 2017), múltiplos marcadores têm sido empregados em conjunto, melhorando a robustez da identificação taxonômica em estudos de *metabarcoding* (Axtner et al., 2019; Lynggaard et al., 2019; Hajibabaei et al., 2019).

Atualmente, a abordagem de *metabarcoding* de amostras ambientais tem sido amplamente utilizada em levantamentos e monitoramentos ao redor do mundo, abrangendo ambientes marinhos, de água doce e terrestres, e permitindo a detecção de espécies difíceis de monitorar por métodos convencionais (Belle et al., 2019; Deiner et al., 2017). Apesar de sua ampla adoção, a técnica enfrenta desafios, como a necessidade de bancos de referência mais abrangentes, uma vez que muitas espécies ainda não estão representadas (Tzafesta et al., 2022). Esse desafio é particularmente significativo na região neotropical, onde a imensa complexidade e riqueza de espécies, somadas à carência de infraestrutura logística e à dificuldade de locomoção, prejudicam os trabalhos de campo e comprometem a ampliação de bancos de referência (Jackman et al., 2021; Taberlet et al., 2012).

Apesar do tamanho corporal relativamente grande em comparação com outros grupos, o estudo de mamíferos na região Neotropical apresenta desafios devido a baixas densidades populacionais e à habilidade desses animais de se esconderem na densa e complexa vegetação (Schipper et al., 2008; Schnell et al., 2012). Embora recebam atenção significativa em estudos de conservação, 13% das espécies de mamíferos ainda são classificadas como Deficientes em Dados (DD) pela IUCN, sendo que 35% delas ocorrem nos Neotrópicos (IUCN, 2024). Tradicionalmente, levantamentos de mamíferos envolvem a coleta e/ou observação de espécimes, demandando trabalhos de campo exaustivos. O uso de câmeras-trap tem reduzido o esforço amostral e contribuído para esses estudos, especialmente em habitats densos como florestas (Glover-Kapfer et al., 2019). No entanto, essas câmeras não permitem distinguir espécies próximas e são mais eficazes para mamíferos terrestres de médio e grande porte, sendo pouco úteis para espécies arborícolas, de pequeno porte (menos de 100 g) e morcegos (Bernard et al., 2013).

O *metabarcoding* de amostras ambientais apresenta eficiência igual ou superior às câmeras-trap, permitindo a detecção de mamíferos terrestres, semiaquáticos, arborícolas, de pequeno, médio ou grande porte, incluindo espécies ameaçadas (Ushio et al., 2017; Allen et al., 2022; Keck et al., 2023). Além de se dispersar no ambiente por meio de fezes, urina, células epidérmicas e outras fontes, o DNA pode transitar entre diferentes níveis tróficos de uma teia alimentar, passando, por exemplo, da presa para o predador ou do hospedeiro para o parasita. Dessa forma, moléculas de DNA de mamíferos e outros vertebrados podem ser encontradas no conteúdo intestinal de invertebrados hematófagos e saprófagos, permitindo a identificação de espécies por meio da abordagem de

metabarcoding de iDNA (DNA derivado de invertebrados) (Calvignac-Spencer et al., 2013; Martínez-de la Puente et al., 2015).

Embora a maioria dos estudos de *metabarcoding* de iDNA tenha sido conduzida em regiões temperadas, reflexo do financiamento limitado de pesquisas em países tropicais (Carvalho et al., 2022), um número crescente de estudos utilizando diversos grupos de invertebrados como fontes de DNA de vertebrados já foi aplicado com sucesso na região Neotropical (por exemplo, Kocher et al., 2017; Rodgers et al., 2017; Lynggaard et al., 2019; Massey et al., 2022; Saranholi et al., 2024). No entanto, os primers frequentemente utilizados nesses estudos foram originalmente desenvolvidos para organismos de outras regiões, o que pode levar à inibição da amplificação devido à divergência entre as sequências dos primers e das espécies-alvo, resultando na não-detecção de espécies (falsos-negativos) (Primmer et al., 1996). Assim, é necessário adaptar os marcadores genéticos à vasta diversidade Neotropical para aumentar a eficiência e a taxa de detecção de espécies (Ficetola et al., 2021; Teixeira et al., 2023).

Diferentes fontes de iDNA podem introduzir diferentes vieses taxonômicos devido a variações na ecologia alimentar e no ciclo de vida dos organismos (Massey et al., 2022). Moscas varejeiras, das famílias Calliphoridae e Sarcophagidae, têm sido frequentemente utilizadas como fontes de iDNA, permitindo a detecção de ampla diversidade de vertebrados, incluindo mamíferos terrestres, voadores e arborícolas (Calvignac-Spencer et al., 2013; Rodgers et al., 2017; Gogarten et al., 2020; Lee et al., 2023). Moscas hematófagas da família Tabanidae (mutucas) também possuem potencial como fontes de iDNA, pois são se alimentam oportunisticamente em uma variedade de hospedeiros, incluindo mamíferos e aves (Kniepert 1980; Vaduva, 2015). Um trabalho recente foi capaz de identificar com sucesso espécies de mamíferos e aves através de mosquitos e moscas de várias famílias (Saranholi et al. 2023), porém as mutucas ainda não foram utilizadas como “amostradoras” de vertebrados.

Dípteros hematófagos e saprófagos possuem sistemas sensoriais complexos e altamente especializados para localizar potenciais hospedeiros que servem como fontes de alimento essenciais à sua sobrevivência e reprodução. Com alta capacidade de dispersão no ambiente (Brown, 2020), proporcionada por seu tamanho e habilidade de voo, esses insetos conseguem alcançar mamíferos furtivos em ambientes tropicais. Diferentemente dos mamíferos, a coleta de dípteros pode ser realizada de forma simples e consistente. Armadilhas de interceptação de voo, como as armadilhas Malaise, requerem baixo esforço amostral e conseguem capturar grande número e diversidade de dípteros (Blahó et al., 2013; Lynggaard, 2019; Skvarla et al., 2021).

O objetivo desta tese foi testar a eficácia do *metabarcoding* de iDNA para o levantamento de espécies de mamíferos na região neotropical, com foco nos biomas brasileiros, de forma que as informações obtidas por meio dessa abordagem possam ser utilizadas em programas de biomonitoramento e estudos biogeográficos, contribuindo para a conservação. No primeiro capítulo, avaliamos os marcadores genéticos e os pares de primers mais utilizados na literatura, selecionando os mais adequados para a identificação taxonômica de mamíferos brasileiros. Também personalizamos um banco de referência, reunindo sequências de mamíferos brasileiros disponíveis em bancos de dados públicos online. No segundo capítulo, validamos o desempenho desses primers e do banco de referência em amostras reais de iDNA, expandindo a aplicação do *metabarcoding* em três biomas brasileiros (Amazônia, Mata Atlântica e Cerrado) e utilizando moscas de três famílias distintas (Calliphoridae, Sarcophagidae e Tabanidae).

2. Capítulo I. Seleção de marcadores genéticos e montagem de banco de dados de referência para *metabarcoding* de mamíferos neotropicais (Artigo submetido à Conservation Genetics Resources)

RESUMO

Os biomas neotropicais são caracterizados por uma alta biodiversidade e por ameaças significativas. Para atender à necessidade urgente de avaliação da biodiversidade nesses ecossistemas ameaçados e amplamente subexplorados, como os do Brasil, técnicas modernas como o *metabarcoding* têm sido empregadas para gerar inventários taxonômicos detalhados. No entanto, a seleção de marcadores genéticos apropriados continua sendo desafiadora, pois diferentes marcadores e pares de primers podem introduzir vieses ao amplificar preferencialmente certos grupos taxonômicos, dependendo de suas sequências nucleotídicas e do banco de referência utilizado. Este estudo avança na pesquisa de biodiversidade ao desenvolver um banco de referência customizado e otimizar primers para o *metabarcoding* de mamíferos brasileiros. Nosso banco de referência, cobrindo 72% das espécies de mamíferos nativas, incluindo espécies ameaçadas, melhora a precisão da identificação taxonômica em todos os biomas brasileiros. Testes *in silico* e *in vitro* foram realizados para avaliar e otimizar os marcadores genéticos e pares de primers mais adequados, comparando sua cobertura taxonômica e resolução. Modificações personalizadas feitas nos primers de COI reduziram incompatibilidades com as sequências de mamíferos brasileiros, aumentando significativamente a cobertura taxonômica e superando os primers comumente utilizados na literatura. Demonstramos que a combinação de COI com RNA12S e RNA16S fornece informações importantes e complementares, melhorando a robustez das atribuições taxonômicas e enfrentando os desafios impostos pela diversidade de mamíferos neotropicais. Essa abordagem multimarcadores oferece uma estratégia confiável para aprimorar a identificação por *metabarcoding* na região neotropical, apoiando esforços de conservação mais eficazes.

Title
Metabarcoding Markers and a Reference Database for Neotropical Mammals

Author information
Barbara R N Chaves (barbarachaves@ufmg.br, https://orcid.org/0009-0008-0977-8974)
Jose Eustaquio Santos-Junior (jrsantos140782@yahoo.com.br, https://orcid.org/0000-0002-7150-3751)
Fabricio R Santos (fsantos@icb.ufmg.br, https://orcid.org/0000-0001-9088-1750)
Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brazil

Abstract

Neotropical biomes are characterized by both high biodiversity and significant threats. To address the urgent need for biodiversity assessment in these threatened and largely understudied ecosystems, like the ones in Brazil, modern techniques like metabarcoding have been employed to generate detailed taxonomic inventories. However, selecting appropriate genetic markers remains challenging, as different markers and primer pairs may introduce biases by preferentially amplifying certain taxonomic groups, depending on their nucleotide sequences and the reference database used. This study advances biodiversity research by developing a customized reference database and optimizing primers for metabarcoding Brazilian mammals. Our database, covering 72% of native species, including endangered ones, enhances taxonomic identification accuracy across all Brazilian biomes. Both *in-silico* and *in-vitro* tests were conducted to evaluate and optimize the most suitable genetic markers and primer pairs, comparing their taxonomic coverage and resolution. Customized modifications made to COI primers reduced mismatches against Brazilian mammal sequences, significantly enhancing taxonomic coverage and outperforming commonly used primers in literature. We demonstrate that combining COI with RNA12S and RNA16S provides important and complementary information, improving the robustness of taxonomic assignments and addressing challenges posed by the Neotropical mammalian diversity. This multi-marker approach offers a reliable strategy for enhancing metabarcoding identification in the Neotropical region, supporting more effective conservation efforts.

Keywords
Metabarcodes, Neotropics, Brazilian Mammals, COI, Reference Database, Molecular Taxonomy

Acknowledgments
Thanks to the financial support of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio), and Universidade Federal de Minas Gerais (UFMG).

Text (total word count: 5305+208)
INTRODUCTION (word count: 1038)

Neotropical biomes are among the most degraded and rapidly disappearing natural environments worldwide, and many of them are considered biodiversity hotspots – some of the richest and most threatened biomes (Myers *et al.* 2000; Jenkins *et al.* 2013). Many neotropical species are becoming locally extinct or disappearing before being described by science, which underscores the urgent need for detailed taxonomic inventories and for collecting essential species information (Thomsen *et al.* 2012). Modern techniques have facilitated and accelerated species identification and research. Over two decades ago, DNA barcoding was proposed, revolutionizing taxonomy research by sequencing short DNA sequences of molecular markers from individual specimens for species identification (Hebert *et al.* 2003). More recently, metabarcoding has evolved as a tool for biodiversity monitoring (Taberlet *et al.* 2012), sequencing even shorter DNA fragments from complex mixtures of multiple specimens, usually using high-throughput sequencing (NGS) techniques.

Currently, metabarcoding is widely employed for biodiversity surveys from environmental samples (eDNA) such as soil and water, or from faeces and stomach contents (e.g., iDNA) (Valentini *et al.* 2009; Alberdi *et al.* 2018). The taxonomic classification of genetic sequences and the resulting species list from metabarcodes provide a foundation for subsequent ecological analyses and biodiversity monitoring (Keck *et al.* 2023). Various metabarcoding protocols are available, generally comprising: i) DNA extraction, ii) amplification of one or more genetic markers using specific primers, iii) DNA sequencing using NGS, and iv) bioinformatic analysis for taxonomic assignment. NGS techniques offer the advantage of identifying virtually all species present in a sample, even at low abundances (Kircher & Kelso 2010).

The standard molecular marker for animal DNA barcoding is a 648 base-pair (bp) region of the mitochondrial gene Cytochrome Oxidase Subunit I (COI) (Hebert *et al.* 2003), however there is no standard molecular marker for metabarcoding. Due to its high taxonomic resolution and extensive database (Hajibabaei 2012; Clarke *et al.* 2017; Andújar *et al.* 2018), 648bp-COI could be considered suitable for metabarcoding. However, it is deemed too large for the degraded and fragmented DNA typically found in eDNA samples (Kent & Norris 2005; Alcaide *et al.* 2009; Kent 2009). Additionally, its high evolutionary rate results in a lack of conserved sites for primer design, which can lead to amplification biases, whereby primers target only part of the eDNA sequences (Kent 2009; Deagle *et al.* 2014; Piper *et al.* 2019), although degenerate primers can be used to expand taxonomic coverage (Kwok *et al.* 1990; Wei *et al.* 2003).

Small fragments of other mitochondrial genes, such as ribosomal genes (12S and 16S), are often used for the amplification of a wide range of taxa, including undescribed species (Green *et al.* 2015). These genes have already provided accurate species-level identifications in some studies worldwide (e.g. Ficetola *et al.* 2010; Riaz *et al.* 2011; Piper *et al.* 2019). However, ribosomal genes generally exhibit significant overlap between intra- and interspecific genetic distances and have a smaller reference database compared to COI, which reduces their resolution (Alberdi *et al.* 2017). Furthermore, single genetic markers can produce biased results, as shorter sequences may lack species-specific mutations (Freeland 2017), and different primer pairs may favour certain species groups over others (Alberdi *et al.* 2017; Piper *et al.* 2019).

Nevertheless, the use of multiple genetic markers can provide complementary information and enhance taxonomic coverage (Corse *et al.* 2019; Hajibabaei *et al.* 2019). Consequently, multiple markers were used to improve the robustness of taxonomic identification in metabarcoding studies (Pompanon *et al.* 2012; Alberdi *et al.* 2017; Axtner *et al.* 2019; Lynggaard *et al.* 2019). A search through previous studies reveals the use of several different molecular markers and primer pairs across metabarcoding studies. Choosing appropriate markers and primer pairs is crucial and should consider: i) taxonomic coverage, i.e., which taxa need to be amplified, ii) taxonomic resolution, i.e., the level of taxonomic discrimination, and iii) the availability of reference databases for taxonomic identification (Townzen *et al.* 2008; Pompanon *et al.* 2012; Reeves *et al.* 2018).

The taxonomic coverage of a molecular marker can be limited when primers used were designed for distantly related species, due to increased sequence divergence, including mutations at priming sites. Mismatches can accumulate, inhibiting amplification and leading to the non-detection of occurring species, resulting in false negatives (Primmer *et al.* 1996). Consequently, the analysis of amplified DNA can provide a distorted community composition, with some species potentially missed due to systematic biases in amplification efficiency (Green *et al.* 2015). This issue is particularly relevant in the Neotropical region, known for its vast genetic and species diversity (Raven *et al.* 2020), and can be exacerbated when primers are designed for organisms from other regions, usually from biomes in the Global North.

The quality of the reference database is also crucial for reliable taxonomic assignments, as it links genetic sequences to established taxonomic classifications. Incomplete databases can lead to false negatives and, when combined with sequence and taxonomic errors, result in misidentification – one of the most critical issues to avoid (Keck *et al.* 2023). To minimize misidentification, geographic filtering of sequences can improve the

database by excluding species known to occur outside the study area. Additionally, harmonizing taxonomy and resolving synonyms are important to prevent taxonomic conflicts that may arise when using large public databases (Grenié *et al.* 2023).

This study aimed to establish technical guidelines for the effective use of metabarcodes for identifying Brazilian mammal species, including recommendations for specific primers and reference databases. To achieve this, we constructed a reliable reference database by collecting sequence records of Brazilian mammalian species from public databases. Additionally, the most frequently used genetic markers and primer pairs in the literature were evaluated to select the most suitable ones for the taxonomic identification of Brazilian mammalian species. Both *in-silico* and *in-vitro* tests were conducted to compare the performance of each genetic marker and primer pair with respect to taxonomic coverage and resolution, ensuring their effectiveness for identifying Brazilian mammal species.

Given the high level of ongoing degradation due to increasing anthropogenic impact and an urgent need for conservation in tropical ecosystems, particularly in Brazil's rich and threatened biomes, it is crucial to optimize biodiversity assessment methods. By refining metabarcoding practices, we hope to contribute to more accurate taxonomic inventories and better-informed conservation strategies for these critical habitats of the Global South.

METHODS (word count: 1261)

Reference Databases Construction

A customized reference database for reliable taxonomic assignments of Brazilian mammals was assembled based on the list of native species from the Brazilian Society of Mammalogy (SBMz) (Abreu *et al*. 2023), supplemented with species listed in Mammal Species of the World (Wilson & Reeder 2005) and records of exotic species in Brazil (Da Rosa *et al*. 2017). This list contains essential species information, including IUCN threat status (2024), ecological niche, and biome of occurrence (Abreu *et al*. 2023; Paglia *et al*. 2012). To address taxonomic inconsistencies among databases (Keck *et al*. 2023), synonyms were included to cross-reference taxonomic systems (SBMz, IUCN, NCBI, BOLD, MSW).

Sequence records from COI, RNA12S, RNA16S, and complete mitogenomes for each species and their synonyms were downloaded from GenBank (Clark *et al*. 2016), using the *traits* package in R (LeBauer *et al*. 2024), and from BOLD (Ratnasingham & Hebert 2007), using the *bold* package in R (Mudalige 2021). Taxonomy for GenBank sequences was assigned using the *taxonomizr* package (Sherrill-Mix 2019) and for BOLD sequences using the *bold* package. To capture intraspecific diversity, up to three sequences per marker were retained for each species. To minimize missing taxa, additional sequences from closely related species from other regions were included, selecting one sequence per species within each native genus. To enable broader taxonomic identification, one complete mitogenome per genus was added for selected animal classes (Aves, Amphibia, Lepidosauria, Actinopteri, and Insecta). All mitogenomes were divided into three parts (0–2,000 bp, 1,550–3,600 bp, and 5,000–9,500 bp) and included as RNA12S, RNA16S, and COI sequences, respectively.

Genetic Markers and Primers Selection

To identify suitable genetic markers for Brazilian mammals, the most commonly reported primer pairs in the literature were compiled. Chemical properties of the primers, including length, GC content, melting temperature, 3' end specificity, and dimerization (self or cross annealing), were analysed using PCR Primer Stats (Stothard 2000) and Multiple Primer Analyzer (ThermoFisher 2024) to evaluate quality and PCR efficiency (Banaganapalli *et al*. 2019). The lengths of candidate genetic markers (amplicon sizes) were verified on references or predicted by Primer-BLAST (Ye *et al*. 2012). As metabarcoding is often applied to degraded samples with short DNA fragments (e.g., soil, water, and faeces/stomach contents) (Freeland 2017), shorter genetic markers (max 300 bp) are preferred.

An ideal marker should cover as many mammal species as possible, prioritizing those in Brazil, and preferably avoiding amplification of invertebrate DNA to ensure effective application on iDNA samples. Specificity at the species level is crucial, especially for ecological and conservation studies, and a comprehensive public database is needed to reduce false negatives or misidentifications. To address these requirements, *in-silico* and *in-vitro* tests were conducted to assess the performance of different markers and primers.

In-silico tests were conducted on the compiled primer pairs using the *primerTree* package in R (Cannon *et al*. 2016), retrieving sequences from GenBank that each primer could amplify. Retrieved sequences were categorized as mammalian, Brazilian mammalian, or dipteran taxa. Primer pairs were evaluated based on their ability to: i) amplify the most mammal species, ii) cover the highest number of mammal families in Brazil, and iii) target the most mammal families overall. Primers amplifying more dipteran than mammal sequences were excluded. Given the more extensive COI database, further optimization of the best COI primer pair was

undertaken to improve mammalian identification accuracy. COI sequences from GenBank, including 130 native Brazilian mammal species across 36 families and 11 orders, were aligned to guide modifications and adjust primers to reduce mismatches against the target sequences. Modified versions of forward (F) and reverse (R) primers were created and retested *in-silico* using *primerTree* to compare their performance with the original primers.

The most effective primers identified *in-silico* were synthesized and tested *in-vitro* to verify amplification efficiency and taxonomic resolution at the species level. Samples morphologically identified and preserved in the DNA/tissue collection of *Centro de Coleções Taxonômicas* (CCT-UFMG) were used as controls, representing Brazilian mammal diversity (Table 1). These samples were incidentally collected over decades and provisionally identified, some accompanied by voucher specimens in CCT-UFMG's mastozoological collection, all benefiting from DNA-based identification for confirmation or correction of taxonomic classifications. All molecular assays performed for *in-vitro* tests were conducted at the Laboratory of Biodiversity and Molecular Evolution of UFMG (LBEM-UFMG).

Table 1. Mammal samples used *in-vitro* to test genetic markers and primers, including provisional taxonomic identifications based on morphology, numbers of sequences available for COI, RNA12S, and RNA16S in the customized reference database, and molecular taxonomy diagnostics using Sanger and Illumina sequencing, according to Fig 1.

Provisional Taxonomy (Morphology)			**Reference database[a]**			***In-vitro* tests**			
Order	**Family**	**Species**	**COI**	**RNA 12S**	**RNA 16S**	**Sanger[b]**	**Illumina[c]**		
						COI	**COI[d]**	**RNA 12S[e]**	**RNA 16S[f]**
Artiodactyla	Delphinidae	*Sotalia fluviatilis*	6	5	5	?	?	☑	no seq
		Sotalia guianensis	5	4	2	no seq	no seq	no seq	no seq
	Iniidae	*Inia geoffrensis*	9	8	8	no seq	NA	NA	NA
Carnivora	Canidae	*Cerdocyon thous*	3	0	0	⚠ *Canis lupus*	no seq	no seq	no seq
	Felidae	*Puma yagouaroundi*	5	5	5	☑	no seq	no seq	☑ pident<95
	Mustelidae	*Lontra longicaudis*	6	6	5	☑	☑	no seq	☑
	Procyonidae	*Nasua nasua*	5	5	5	☑	☑	no seq	no seq
Chiroptera	Emballonuridae	*Rhynchonycteris naso*	3	2	0	☑	☑	☑	no seq
	Molossidae	*Molossops temminckii*	3	0	0	☑	☑	⚠ *Noctilio leporinus*	no seq
	Mormoopidae	*Pteronotus sp.*	3	0	0	☑ *P. gymnonotus*	no seq	no seq	no seq
	Noctilionidae	*Noctilio albiventris*	3	3	0	☑	☑	⚠ *N. leporinus*	⚠ Anura
	Phyllostomidae	*Artibeus lituratus*	6	6	6	☑	☑ pident<95	no seq	☑
		Desmodus rotundus	6	6	3	☑	no seq	no seq	☑ pident<95
		Glossophaga soricina	6	6	3	☑	NA	NA	NA
		Phyllostomus discolor	6	6	3	☑	NA	NA	NA
		Phyllostomus hastatus	3	3	0	no seq	NA	NA	NA
	Vespertilionidae	*Eptesicus brasiliensis*	3	0	0	?	no seq	?	no seq
Cingulata	Dasypodidae	*Dasypus novemcinctus*	6	6	6	⚠ *D. septemcinctus*	no seq	no seq	no seq
Didelphimorphia	Didelphidae	*Didelphis marsupialis*	5	5	4	⚠ *D. albiventris*	☑	☑	☑
Lagomorpha	Leporidae	*Sylvilagus brasiliensis*	3	3	0	no seq	NA	NA	NA
Pilosa	Bradypodidae	*Bradypus torquatus*	6	5	6	☑ pident<95	☑	☑	⊘ pident<95
	Cyclopedidae	*Cyclopes didactylus*	6	6	6	☑ pident<95	no seq	no seq	no seq
	Myrmecophagidae	*Myrmecophaga tridactyla*	4	6	6	☑	☑ pident<95	no seq	☑
		Tamandua tetradactyla	6	6	6	no seq	no seq	☑	no seq
Primates	Atelidae	*Alouatta fusca*	4	3	3	☑	☑	⚠ *Sapajus nigritus*	☑
	Cebidae	*Brachyteles arachnoides*	3	5	4	no seq	NA	NA	NA
		Callithrix penicillata	4	3	3	☑	no seq	☑ pident<95	no seq
		Mico rondoni	0	0	0	⚠ *M. argentatus*	no seq	no seq	⊘
	Pitheciidae	*Callicebus dubius*	3	3	3	⚠	no seq	no seq	⚠

Provisional Taxonomy (Morphology)			Reference database[a]			*In-vitro* tests			
						Plecturocebus caligatus			*P. moloch*
Rodentia	Cricetidae	*Necromys lasiurus*	3	0	0	⚠ *Oligoryzomys fornesi*	no seq	⚠ *Akodon montensis*	⚠ *Akodon montensis*
	Cuniculidae	*Cuniculus paca*	5	5	2	no seq	☑	☑	☑
Sirenia	Trichechidae	*Trichechus inunguis*	1	1	1	⚠ *T. manatus*	⚠ *Artibeus lituratus*	☑	☑
		Trichechus manatus	6	6	6	no seq	NA	NA	NA

a: Number of sequences available for each genetic marker, supplemented with mitogenome sequences; **b**: Sanger sequencing using different combinations of F and R modified primers from Lee *et al.* (2015) followed by *blastn* (Altschul *et al.* 1990) to identify mammalian samples; **c**: two-step PCR protocol for Illumina sequencing (Ushio *et al.* 2017, Chen *et al.* 2021) followed by PIMBA pipeline (Oliveira *et al.* 2021) to identify mammalian samples; **d**: different combinations of F and R modified primers from Lee *et al.* (2015); **e**: original primers from Ushio *et al.* (2017); **f**: primers modified from Haile *et al.* (2009); *NA*: samples not used for Illumina sequencing.

The efficiency of the modified COI primers was evaluated through *in-vitro* amplification and Sanger sequencing of 33 mammal species across 24 families and 10 orders, following standard protocols (Chaves *et al*. 2015). Consensus sequences were assembled in UGENE (Okonechnikov *et al*. 2012) and matched against our customized database using *blastn* (Altschul *et al*. 1990) to verify correct COI gene region amplification and accurate species identification. For unidentified samples, an additional *blastn* run was conducted against the full GenBank nucleotide collection, retaining only the best match per sequence.

Taxonomic resolution of the genetic markers was evaluated using a two-step PCR protocol for Illumina library preparation, adapted from Ushio *et al*. (2017) and Chen *et al*. (2021). In the first PCR, MiSeq sequencing primers were added as 5' tails to each marker primer for *in-vitro* amplification of 26 mammal species across 22 families and 9 orders. In the second PCR, 7-base tag identifiers (Hamady *et al*. 2008) and Illumina adapters (P5 or P7) were added to first-step amplicons. Primer modifications, PCR conditions and recipes are detailed in Online Resource 1. After each PCR, amplified products were visualized in 2% agarose gel electrophoresis and purified with magnetic beads (AMPure XP). Second-step products were normalized, pooled, and sequenced on the Illumina MiSeq platform at LG-UFMG using MiSeq Reagent Kit v2 (Illumina) with 25% PhiX.

Illumina reads were processed using the PIMBA pipeline (Oliveira *et al*. 2021) for demultiplexing, quality filtering, OTU clustering, error correction, and taxonomic assignment. OTUs under 100 bp, over 260 bp, or deviating by more than 19 bp from the expected marker size were excluded, as well as OTUs with fewer than 10 reads or representing less than 1% of total sample reads. Our customized database was used for taxonomic assignment, verifying the expected marker and species. OTUs without identification, unexpected matches, or with less than 95% identity were reanalysed via blastn against the full GenBank database, retaining only the best match. Each taxonomic assignment was reviewed with a diagnostic tree of yes/no questions, comparing genetic and morphological identifications of control samples (Fig 1). Possible outcomes included: species OK, wrong species, genus OK, unidentified, or contamination. Primer pairs and genetic markers were evaluated for their amplification efficiency, taxonomic resolution, and database coverage based on: i) the number of amplified control samples, ii) accurate species-level identifications, and iii) database completeness.

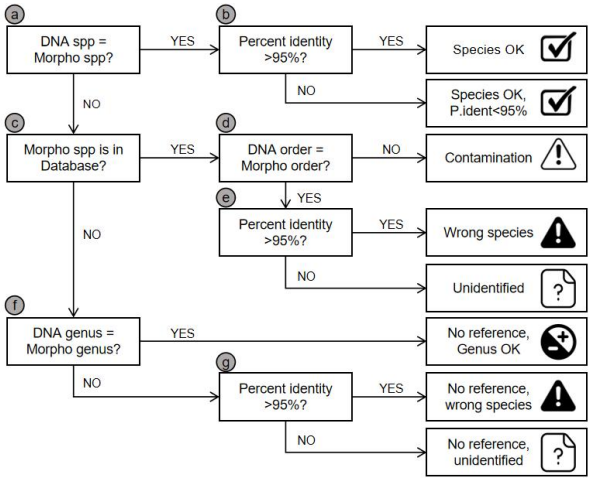

Fig 1 Diagnostic tree with seven steps of questions of "YES" and "NO", labelled from a to g, used to assess DNA-based species identification compared to morphological identification. See Online Resource 1 for a detailed description of steps.

RESULTS AND DISCUSSION (word count: 2819)

Reference Databases Construction

A total of 830 mammal species, including 270 genera, 55 families, and 11 orders potentially present in Brazil, were compiled from the SBMz list (Abreu *et al*. 2023) and other sources (Wilson & Reeder 2005; Da Rosa *et al*. 2017). This list includes 145 species classified as threatened or near threatened (IUCN 2024), 21 exotic species, and 312 synonyms, covering endemic and non-endemic species across all Brazilian biomes (Amazon, Cerrado, Atlantic Forest, Caatinga, Pampa, Pantanal, and Marine) and a range of ecological niches (aquatic, terrestrial, arboreal, scansorial, fossorial, and flying mammals). The complete species list is available in Online Resource 2.

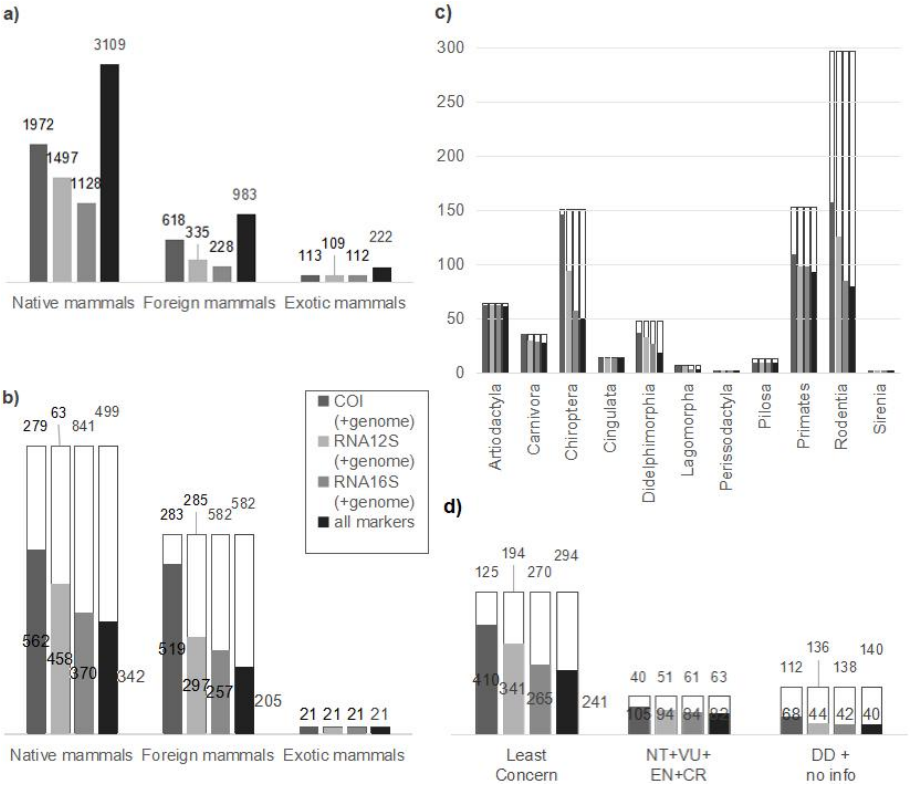

Fig 2 Composition of the customized database for Brazilian mammals, assembled with native species (Abreu *et al*. 2023; Wilson & Reeder 2005), exotic species in Brazil (Da Rosa *et al*. 2017), and related foreign species. The figure displays: a) sequence counts for COI, RNA12S, RNA16S, and mitogenomes from GenBank and BOLD; b) counts of native, foreign, and exotic mammal species with (solid colours) and without (transparent colours) sequences for each marker; c) species count with/without sequences per order; d) species count with/without sequences by IUCN threat categories (NT: Near Threatened, VU: Vulnerable, EN: Endangered, CR: Critically Endangered, DD: Data Deficient).

Starting with this list, our customized database included 7,885 sequences: 6,949 from GenBank and 936 from BOLD. Of these, 3,109 sequences correspond to 841 native Brazilian species, with 744 from mitogenomes, 1,228 from COI, 753 from RNA12S, and 384 from RNA16S. After splitting mitogenome sequences into three segments as RNA12S, RNA16S, and COI, the database contained 1,972 COI sequences for 562 species (67% of Brazilian species), 1,497 RNA12S sequences for 458 species (54%), and 1,128 RNA16S sequences for 370 species (44%) (Fig 2a). No sequences were available for 237 Brazilian species (28%), while 342 species (41%) had sequences for all three markers (Fig 2b), many with multiple sequences per species, representing intraspecific diversity. Additionally, the database includes 222 sequences from 21 exotic species (including *Homo sapiens*), all with sequences for all three markers.

COI sequences were more widely available across mammalian orders than RNA12S and RNA16S, making COI a desirable marker for accurate metabarcoding (Fig 2c). All few and conspicuous species in Cingulata, Perissodactyla, and Sirenia had sequences for all markers, while Rodentia, Didelphimorphia, and Chiroptera were less complete, reflecting their high diversity and sampling challenges. Most threatened species belong to Primates, Rodentia, and Artiodactyla, with 57% having sequences for all markers (Fig 2d), indicating a historical focus on sequencing endangered species. In contrast, species data deficient or with unknown threat status, mainly from Rodentia, Chiroptera, and Didelphimorphia, had only 22% of species represented in all markers,

emphasizing challenges in small mammal studies. Across all threat levels, COI was the most represented marker, supporting its use in identifying threatened species.

To broaden taxonomic scope, 983 sequences from foreign species within the same genus as native species were added (Fig 2a), reducing the number of native genera without sequences from 101 (40%) to 31 (12%). While these sequences cannot fully prevent misidentification, they may offer insights into the taxonomy of native species that are not represented in the database and would otherwise result in false negatives or no-identifications. Additionally, 3,571 sequences from other animal classes (one per genus) were included: 797 from Aves, 261 from Amphibia, 237 from Lepidosauria, 2,029 from Actinopteri, and 247 from Insecta. While non-mammalian sequences do not guarantee precise identification, they help detect contaminants or sequences from other animal classes. The full database is available in Online Resource 3.

Genetic Markers and Primers Selection

A total of 24 primers commonly used for mammalian identification were compiled, including three pairs for COI, two for Cyt-b, five for RNA12S, and three for RNA16S (Table 2). After chemical evaluation, only four primers (COI_short f, 12S-V5-R1, 12S-V5-F2, and Riaz16S1 R) had ideal properties, while the others exhibited features that potentially reduce PCR efficiency: excessive or insufficient primer length (three primers), high or low GC content (15 primers), unsuitable melting temperatures (three primers), ineffective GC clamps (three primers), and self-dimerization (two primers) (details in Online Resource 4). This suggests that many of the primers currently used for mammalian identification are suboptimal, lacking critical characteristics for optimal performance in PCR reactions (Banaganapalli *et al.* 2019). Additionally, three primer pairs (TOWNZEN COI, KOCHER, and TOWNZEN Cytb) generated amplicons longer than 300 bp, unsuitable for metabarcoding (Banaganapalli *et al*. 2019).

Table 2. Compilation of primers from literature used for amplification of different genetic markers of mammalian species, with respective chemical evaluations (Banaganapalli *et al.* 2019), predicted amplicon size (Ye *et al.* 2012), and *in-silico* amplifications of mammalian and dipteran species (Cannon *et al.* 2016).

Marker	Primer pair	Primer reference	F Primer	R Primer	Amplicon size (bp)	*In-silico* PCR			
						Mam spp	Mam Fam (Overall)	Mam Fam (Brazil)[a]	Diptera spp
COI	LEE[4th]	Lee *et al.* 2015	Uni-Mini-bar F	RonPing R	205	224	37	11 (22%)	0
	MEUSNIER[b]	Meusnier *et al.* 2008	Uni-Mini-bar F	Uni-Mini-bar R	130	43	13	7 (14%)	158
	TOWNZEN COI	Townzen *et al.* 2008	COI_short f	COI_short r	324	27	12	7 (14%)	0
	CHAVES[c]	This study	coiMam (F1+F2+F3+F4)	coiMam (R1+R2+R3)	205	657	118	24 (49%)	4
Cytb	KOCHER	Kocher *et al.* 1989	Kocher CytB-fw	Kocher CytB-rv	306	4	1	1 (2%)	0
	TOWNZEN Cytb	Townzen *et al.* 2008	Townzen Cytb F	Townzen Cytb R	450	21	7	3 (6%)	0
RNA12S	JI	Ji *et al.* 2020	12S-V5-F1	Ji2020-R	82 - 150	110	19	8 (16%)	0
	RIAZ 12S 1	Riaz *et al.* 2011	12S-V5-F1	12S-V5-R1	105	103	26	13 (27%)	0
	RIAZ 12S 2	Riaz *et al.* 2011	12S-V5-F2	12S-V5-R2	98	95	24	12 (25%)	0
	TABERLET[3th]	Taberlet *et al.* 2018	Mamm01 F	Mamm01 R	60	234	38	13 (27%)	0
	USHIO[5th]	Ushio *et al.* 2017	MiMammal-U F	MiMammal-U R	171	132	31	14 (29%)	0
RNA16S	HAILE[2nd]	Haile *et al.* 2009	16Smam3	16Smam4	78	406	38	14 (29%)	0
	HAILE 5[1st]	Modified from Haile *et al.* 2009	16Smam3c	16Smam4	78	522	48	17 (35%)	0
	RIAZ 16S	Riaz *et al.* 2011	Riaz16S1 R	Riaz16S1 F	58	88	24	8 (16%)	0
	TAYLOR	Taylor 1996	16Smam1	16Smam2	300	82	25	10 (20%)	0

1st to 5th: best primer pairs selected due to in-silico *taxonomic coverage and specificity.* ***a****: Number of families of Brazilian mammals according to Brazilian Society of Mammalogy (Abreu et al. 2022).* ***b****: Primer pair discarded for amplifying Diptera species.* ***c****: Modified primers from Lee et al. (2015), based on an alignment of mammalian species native to Brazil (see Table 3).*

Despite potential limitations, *in-silico* tests were performed on all 24 primers, which retrieved numerous sequences, which were taxonomically identified, quantified, and categorized according to their respective taxonomic groups. The number of species and families obtained from each primer pair were summarized in

Table 2. The Cyt-b primers (KOCHER and TOWNZEN-Cytb) were the least effective, amplifying fewer mammalian species. The HAILE primer pair, originally developed to amplify a 78 bp of RNA16S and with a small modification in the forward primer (changing the IUPAC Y nucleotide at the 3' end to C), showed the best performance, retrieving sequences for 522 mammalian species from 48 families, 17 of which occur in Brazil (35% of Brazilian mammalian families). Approximately 100 previous studies have used the 12S-V5 primers (Riaz *et al*. 2011), but these did not perform as effectively in the context of this study. Instead, two other RNA12S primer pairs—TABERLET and USHIO—retrieved more species and families, aligning better with metabarcoding needs (Taberlet *et al*. 2018; Ushio *et al*. 2017).

The MEUSNIER primer pair was discarded due to its ability to amplify more dipteran than mammalian species. Originally developed for a short region of the COI gene, these primers can amplify a wide range of taxa, including mammals, fish, birds, insects, plants, fungi, and macroalgae (Meusnier *et al.* 2008). Although their universality makes them valuable for identifying diverse environmental samples, we aimed to avoid amplifying insect species to ensure the primers' suitability for iDNA samples.

Among the COI primers, the best-performing pair was LEE, which was originally adapted from Meusnier's universal primers for more specific amplification of mammals (Lee *et al.* 2015). In our *in-silico* tests, this primer pair retrieved 224 mammalian species from 37 families, 11 of which occurring in Brazil (22% of all Brazilian mammalian families). To further optimize the performance of LEE primer pair, 18 different versions of primer F (Uni-Mini-bar) and 16 versions of primer R (RonPing) were made based on an alignment of mammalian species native to Brazil and manual modifications in their nucleotide sequences to reduce mismatches between primers and target sequences (see more in Online Resource 4). All modified primers were tested *in-silico* and the results of the seven best (four F primers and three R primers) showed significant improvement compared to the original primers (Table 2 and Table 3), achieving the best performance among all primers analysed.

Table 3. Original and the best modified versions of COI primers from Lee *et al.* (2015), with respective sequences, chemical evaluation (Banaganapalli *et al.* 2019), *in-silico* amplifications of mammalian and dipteran species (Cannon *et al.* 2016), and *in-vitro* amplifications and sequencing of mammalian species.

Primer type	Primer name	Primer reference	5'-3' sequence	*In-silico* tests[a]				*In-vitro* tests	
				Mam spp	Mam Fam (Overal)	Mam Fam (Brazil)[b]	Diptera spp	Amplified spp	Sequenced spp[c]
Forward	Uni-Mini-bar F	Lee *et al.* 2015	TCCACTAATCACAARGATATTGGTAC	224	37	11	0	-	-
	coiMam-F1	This study	TCAACAAACCAYAAAGAYATTGGTAC	421	71	23	4	24	22
	coiMam-F2	This study	TCCACAAAYCAYAAGGACATTGGCAC	340	61	22	2	17	15
	coiMam-F3	This study	TCAACAAAYCAYAAAGACATTGGTAC	421	64	22	2	25	20
	coiMam-F4	This study	TCAACYAACCACAAAGACATYGGAAC	399	53	19	0	22	18
Reverse	RonPing R	Lee *et al.* 2015	TATCAGGGGCTCCGATTAT	224	37	11	0	-	-
	coiMam-R1	This study	ATRTCRGGGGCTCCAATTAT	657	118	24	3	34	29
	coiMam-R2	This study	ATRTCTGGRGCACCAATTAT	312	51	20	1	21	19
	coiMam-R3	This study	ATRTCRGGTGCTCCAATTAT	363	62	19	2	23	20

***a**: Maximum values between combinations of the primer with corresponding concurrent primers (F or R). **b**: Families of Brazilian mammals according to Brazilian Society of Mammalogy (Abreu et al. 2022). **c**: Species sequenced using Sanger technique.*

All seven modified COI primers amplified mammalian species *in-vitro* (Table 3). Although no single pair amplified all samples, different F-R combinations amplified all 33 tested mammals, with 25 generating consensus sequences. The best combination, coiMam-F1 and coiMam-R1, sequenced most samples, with additional support from four other combinations. The 205 bp COI-based identifications confirmed taxonomic assignments of 15 samples, corrected seven, and accurately identified 22 species from 19 families across eight orders. Three of the sequenced samples remained unidentified, though one (M01543, *Mico rondoni*) reached the correct genus despite lacking COI sequences in the reference database.

Using a two-step PCR protocol for Illumina sequencing (Ushio *et al*. 2017; Chen *et al*. 2021) and PIMBA pipeline (Oliveira *et al*. 2021), mammalian species were amplified, sequenced, and identified *in-vitro* across three markers: i) COI (205 bp), ii) RNA12S (171 bp), and iii) RNA16S (78 bp). COI performed best, identifying most mammalian samples with different F-R combinations (Table 1, Fig 3). Although no single primer pair amplified and identified all samples, combined they correctly identified 17 of 26 samples, spanning 11 families and eight orders. Notably, one sample not sequenced by Sanger for COI (*Cuniculus paca*, M01538) was

successfully sequenced via NGS with the same marker. However, 11 other samples sequenced by Sanger were not sequenced via NGS, possibly due to decreased efficiency in the two-step PCR protocol.

Some misidentifications occurred with RNA12S and RNA16S (Table 1, Fig 3). For instance, *Noctilio albiventris* was misidentified as *N. leporinus* and *Alouatta fusca* as *Sapajus nigritus*, likely due to close genetic distances in ribosomal genes. RNA12S and RNA16S also lacked reference sequences for many control samples, leading to unidentified or misidentified sequences when applying a 95% identity cut-off, except for one sample reaching the correct genus (M01543, supposedly *Mico rondoni*). Nevertheless, COI performance was improved when combined with RNA16S and RNA12S, demonstrating that multiple genetic markers can enhance taxonomic accuracy.

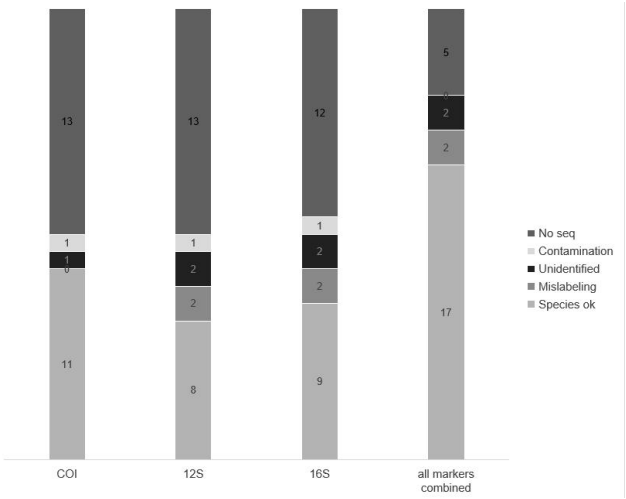

Fig 3 Summary of *in-vitro* tests of selected primers using two-step PCR protocol for Illumina sequencing (Ushio *et al.* 2017, Chen *et al.* 2021) followed by PIMBA pipeline (Oliveira *et al.* 2021) to identify mammalian samples through three different genetic markers: COI (different combinations of F and R modified primers from Lee *et al.* [2015]); RNA12S (original primers from Ushio *et al.* [2017]); and RNA16S (primers modified from Haile *et al.* [2009])

Metabarcoding Performance

Our customized database encompasses 72% of native mammal species and 87% of genera from all orders in Brazil, alongside all exotic species, thereby facilitating the identification of mammals found in aquatic, terrestrial, arboreal, scansorial, fossorial, and flying habitats. By linking genetic sequences to a recognized taxonomic classification (i.e. Abreu *et al.* 2023), this database enhanced the accuracy of species identification, which represents a significant advancement in biodiversity research and conservation. Although further expansion is needed, this represents a positive step toward generating taxonomic inventories and acquiring species information for the diverse and threatened Brazilian biomes. The use of our database in Brazilian inventory surveys may reduce misidentification, as the geographic filtering excludes species known to occur only outside the study area (Grenié *et al.* 2023). If necessary, the database also allows for filtering sequences according to the specific biome under study, ensuring more accurate identifications.

Despite being fairly comprehensive, our customized database has some gaps, particularly among orders Rodentia, Didelphimorphia, and Chiroptera—the same groups that contain most species classified as data deficient or lacking threat information according to the IUCN (2024). This reflects the lack of available sequences in public databases, largely due to high species diversity, low population densities, limitations in trapping methods, and reduced research and conservation efforts for small mammals (Pacheco *et al.* 2013; Santos-Filho *et al.* 2015; Stephenson 2017). This incompleteness can hinder species identification in these orders, potentially leading to false negatives or misidentification (Keck *et al.* 2023). Efforts to increase the sequencing of genetic markers such as COI, RNA12S, and RNA16S in these groups would greatly benefit the surveying of hard-to-sample species.

Our careful analysis of available primers for metabarcoding revealed that many of the primers are suboptimal, lacking characteristics necessary for adequate performance in PCR reactions (Banaganapalli *et al.* 2019) and/or generating too long amplicons, likely due to the challenge of finding stable and ideal regions for primer annealing within genetic marker sequences. The *in-silico* tests allowed to characterize the best-performing primer pairs among the compiled options that were, coincidentally, originally developed or specifically adapted

for mammals: i) a 78 bp region of RNA16S (modified from Haile *et al.* 2009), ii) a 60 bp region of RNA12S (Taberlet *et al.* 2018), iii) a 205 bp region of the COI gene (Lee *et al.* 2015), and iv) a 171 bp region of RNA12S (Ushio *et al.* 2017). These primers amplified the greatest number of mammalian species, particularly those found in Brazil, while avoiding the amplification of insect species, ensuring their potential use in iDNA samples.

Despite the high evolutionary rate and resulting variability in primer-binding sites, COI sequences remain valuable for biodiversity assessments due to their species-level taxonomic resolution and the availability of well-developed databases (Hebert *et al.* 2003; Hajibabaei 2012; Clarke *et al.* 2017; Andújar *et al.* 2018). In fact, our customized database contained a greater abundance of COI sequences across all mammalian orders and threat categories compared to RNA12S and RNA16S. However, some of the primer pairs analysed for RNA12S and RNA16S exhibited better *in-silico* taxonomic coverage than the best COI pair. Our taxon-oriented modifications of the COI primers reduced mismatches against sequences of mammalian species native to Brazil, resulting in significant improvement in their *in-silico* taxonomic coverage.

Although the high variability in primer-binding sites prevented a single primer pair from amplifying COI sequences of all Brazilian mammals—even with degenerate nucleotides—combining different versions of forward (F) and reverse (R) primers ultimately outperformed all other primers analysed *in-silico*. Using coiMam-F1 and coiMam-R1 along with other F-R combinations of modified primers, the expected 205 bp of COI was also successfully amplified from mammals *in-vitro*, accurately identifying two-thirds of the 33 species tested, representing 19 families and 8 orders found in Brazil. This confirms that designing or adapting primers specifically for species from the Neotropical region can enhance their performance. In fact, the results suggest that different combinations of our newly designed COI primers could also be used to identify mammals from other regions of the world with greater efficiency.

Among the best genetic markers identified by *in-silico* tests (Online Resource 5), three were also tested *in-vitro* (Online Resource 6) using a two-step PCR protocol for Illumina sequencing: i) 205 bp of COI, using improved primers modified from Lee *et al.* (2015), ii) 171 bp of RNA12S, using original primers from Ushio *et al.* (2017), and iii) 78 bp of RNA16S, using a modified primer pair from Haile *et al.* (2009). Different F-R combinations of COI primers amplified, sequenced and allowed the accurate identification of more samples compared to RNA12S and RNA16S, making COI the marker with better taxonomic resolution. Additionally, some samples were either unidentified or misidentified by the ribosomal markers (RNA12S and RNA16S), possibly due to the larger incompleteness of their reference databases and the commonly observed overlap between intra- and interspecific genetic distances (Alberdi *et al.* 2017; Kannan *et al.* 2020), where the variability among closely related species can be minimal, making it challenging to distinguish between them.

These results suggest COI as the most accurate metabarcoding identification of mammals, but for certain species, an accurate identification was only achieved using also RNA16S and/or RNA12S. While it was not possible to amplify all samples with a single marker, two-thirds of the 26 species tested—representing 11 families and eight orders—were accurately identified using a combination of all three markers. Thus, we recommend combining COI with ribosomal genes to acquire complementary information and improve the robustness of taxonomic identification in metabarcoding studies of mammals, especially in the highly genetically and biodiverse Neotropical region. Our findings are consistent with previous studies that have also suggested a multi-marker approach to enhance taxonomic coverage in metabarcoding studies (Pompanon *et al.* 2012; Alberdi *et al.* 2017; Axtner *et al.* 2019; Lynggaard *et al.* 2019).

When comparing the two sequencing approaches used for mammalian samples in this study, a notable discrepancy in success was observed. Despite using the same COI primers, half of the samples successfully sequenced and identified through the Sanger technique were not successfully sequenced using the two-step PCR protocol for Illumina. This protocol was chosen because, by introducing indexes only in the second step, it offers primer flexibility and reduces template-specific biases, enabling more uniform amplification and a more concentrated library for sequencing (O'Donnell *et al.* 2016; Chen *et al.* 2021). However, it also presents certain disadvantages, such as an increased risk of contamination (Seitz *et al.* 2015) and a potential impact on amplification efficiency. This occurs because the addition of tail sequences to primers introduces extra nucleotides that can create mismatches with the target sequence, significantly reducing both amplification efficiency and yield (Elbrecht *et al.* 2018; Marquina *et al.* 2018). This issue can be exacerbated when primers are highly degenerate, as the added complexity can further decrease amplification success (Kumar *et al.* 2022).

CONCLUSION (word count: 187)

In this study, we selected and optimized specific primers and customized a reference database for the effective metabarcoding of Brazilian mammals. The development of this database for mammalian identification across Brazilian biomes represents a significant advancement in biodiversity research and conservation. With 72% of

known native species covered, including endangered species, our customized database holds great potential to reduce misidentification in metabarcoding studies in Brazil. Our optimized COI primers improved the accuracy and efficiency of species identification, particularly for mammals in the Neotropical region, by enhancing amplification success and expanding taxonomic coverage for metabarcoding applications. Notably, these newly designed primers could also be applied to mammalian identification in other regions of the world with greater efficiency. In both *in-silico* and *in-vitro* comparisons, COI emerged as the preferred genetic marker due to its extensive reference database and superior taxonomic resolution. However, the combined use with ribosomal sequences was essential for providing complementary information and improving the robustness of taxonomic identification. These findings highlight the effectiveness of a multi-marker approach in addressing the challenges of genetic variability in the Neotropical region and strengthening the accuracy of mammalian metabarcoding studies.

REFERENCES

Abreu, E. F., Casali, D., Costa-Araújo, R., Garbino, G. S. T., Libardi, G. S., Loretto, D., Loss, A. C., Marmontel, M., Moras, L. M., Nascimento, M. C., Oliveira, M. L., Pavan, S. E., & Tirelli, F. P. (2023). Lista de Mamíferos do Brasil (2023-1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10428436

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2017). Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution, 9(1), 134-147. https://doi.org/10.1111/2041-210X.12849

Alcaide, M., Rico, C., Ruiz, S., Soriguer, R., Munoz, J., & Figuerola, J. (2009). Disentangling vector-borne transmission networks: a universal DNA barcoding method to identify vertebrate hosts from arthropod bloodmeals. PloS one, 4(9), e7092. https://doi.org/10.1371/journal.pone.0007092

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andújar, C., Arribas, P., Gray, C., Bruce, C., Woodward, G., Yu, D. W., & Vogler, A. P. (2018). Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. Molecular Ecology, 27(1), 146-166. https://doi.org/10.1111/mec.14410

Axtner, J., Crampton-Platt, A., Hörig, L. A., Mohamed, A., Xu, C. C., Yu, D. W., & Wilting, A. (2019). An efficient and robust laboratory workflow and tetrapod database for larger scale environmental DNA studies. *GigaScience*, 8(4), giz029. https://doi.org/10.1093/gigascience/giz029

Banaganapalli, B., Shaik, N. A., Rashidi, O. M., Jamalalail, B., Bahattab, R., Bokhari, H. A., ... & Elango, R. (2019). In-silico PCR. In Essentials of Bioinformatics, Volume I (pp. 355-371). *Springer*, Cham.

Cannon, M. V., Hester, J., Shalkhauser, A., Chan, E. R., Logue, K., Small, S. T., & Serre, D. (2016). In-silico assessment of primers for eDNA studies using PrimerTree and application to characterize the biodiversity surrounding the Cuyahoga River. *Scientific reports*, 6(1), 1-11. https://doi.org/10.1038/srep22908

Chaves, B. R., Chaves, A. V., Nascimento, A. C., Chevitarese, J., Vasconcelos, M. F., Santos, F. R. (2015). Barcoding Neotropical birds: assessing the impact of nonmonophyly in a highly diverse group. Molecular Ecology Resources, 15, 921-931

Chen, K. H., Longley, R., Bonito, G., & Liao, H. L. (2021). A two-step PCR protocol enabling flexible primer choice and high sequencing yield for Illumina MiSeq meta-barcoding. *Agronomy*, 11(7), 1274. https://doi.org/10.3390/agronomy11071274

Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2016). GenBank. Nucleic acids research, 44(D1), D67-D72. https://doi.org/10.1093/nar/gkv1276

Clarke, L. J., Beard, J. M., Swadling, K. M., & Deagle, B. E. (2017). Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and evolution, 7(3), 873-883. https://doi.org/10.1002/ece3.2667

Corse, E., Tougard, C., Archambaud-Suard, G., Agnèse, J. F., Messu Mandeng, F. D., Bilong Bilong, C. F., ... & Dubut, V. (2019). One-locus-several-primers: A strategy to improve the taxonomic and haplotypic coverage in diet metabarcoding studies. Ecology and Evolution, 9(8), 4603-4620. https://doi.org/10.1002/ece3.5063

Da Rosa, C. A., de Almeida Curi, N. H., Puertas, F., & Passamani, M. (2017). Alien terrestrial mammals in Brazil: current status and management. Biological Invasions, 19(7), 2101-2123. https://doi.org/10.1007/s10530-017-1423-3

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biology letters, 10(9), 20140562. https://doi.org/10.1098/rsbl.2014.0562

Elbrecht, V., Hebert, P. D. N., & Steinke, D. (2018). Slippage of degenerate primers can cause variation in amplicon length. Scientific Reports, 8(1). https://doi.org/10.1038/s41598-018-29364-z

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., ... & Pompanon, F. (2010). An in-silico approach for the evaluation of DNA barcodes. BMC genomics, 11, 1-10. https://doi.org/10.1186/1471-2164-11-434

Freeland, J. R. (2017). The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA. Genome, 60(4), 358-374. https://doi.org/10.1139/gen-2016-0100

Green, S. J., Venkatramanan, R., & Naqib, A. (2015). Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches. PloS one, 10(5), e0128122. https://doi.org/10.1371/journal.pone.0128122

Grenié, M., Berti, E., Carvajal-Quintero, J., Dädlow, G. M. L., Sagouis, A., & Winter, M. (2023). Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices. Methods in Ecology and Evolution, 14(1), 12-25. https://doi.org/10.1111/2041-210X.13802

Haile, J., Froese, D. G., MacPhee, R. D., Roberts, R. G., Arnold, L. J., Reyes, A. V., ... & Willerslev, E. (2009). Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences*, 106(52), 22352-22357. https://doi.org/10.1073/pnas.0912510106

Hajibabaei, M. (2012). The golden age of DNA metasystematics. *Trends in genetics*, 28(11), 535-537. https://doi.org/10.1016/j.tig.2012.08.001

Hajibabaei, M., Porter, T. M., Wright, M., & Rudar, J. (2019). COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLoS One, 14(9), e0220953. https://doi.org/10.1371/journal.pone.0220953

Hamady, M., Walker, J., Harris, J. *et al.* Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods 5, 235–237 (2008). https://doi.org/10.1038/nmeth.1184

Hebert, P. D., Ratnasingham, S., & De Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(suppl_1), S96-S99. https://doi.org/10.1098/rsbl.2003.0025

IUCN. 2024. The IUCN Red List of Threatened Species. Version 2024-1. https://www.iucnredlist.org. Accessed on 18/09/2024.

Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences*, 110(28), E2602-E2610. https://doi.org/10.1073/pnas.1302251110

Ji, Y., Baker, C. C., Popescu, V. D., Wang, J., Wu, C., Wang, Z., ... & Yu, D. W. (2020). Measuring protected-area outcomes with leech iDNA: large-scale quantification of vertebrate biodiversity in Ailaoshan reserve. BioRxiv, 2020-02. https://doi.org/10.1101/2020.02.10.941336

Kannan, A., Rao, S. R., Ratnayeke, S., & Yow, Y. (2020). The efficiency of universal mitochondrial DNA barcodes for species discrimination of *Pomacea canaliculata* and *Pomacea maculata*. PeerJ, 8, e8755. https://doi.org/10.7717/peerj.8755

Keck, F., Couton, M., & Altermatt, F. (2023). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Molecular Ecology Resources, 23(4), 742-755. https://doi.org/10.1111/1755-0998.13746

Kent, R. J. (2009). Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. Molecular ecology resources, 9(1), 4-18. https://doi.org/10.1111/j.1755-0998.2008.02469.x

Kent, R. J., & Norris, D. E. (2005). Identification of mammalian blood meals in mosquitoes by a multiplexed polymerase chain reaction targeting cytochrome B. *The American journal of tropical medicine and hygiene*, 73(2), 336-342.

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing–concepts and limitations. *Bioessays*, 32(6), 524-536. https://doi.org/10.1002/bies.200900181

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences, 86(16), 6196-6200. https://doi.org/10.1073/pnas.86.16.6196

Kumar, G., Reaume, A. M., Farrell, E., & Gaither, M. R. (2022). Comparing edna metabarcoding primers for assessing fish communities in a biodiverse estuary. Plos One, 17(6), e0266720. https://doi.org/10.1371/journal.pone.0266720

Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C., & Sninsky, J. J. (1990). Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucleic acids research, 18(4), 999-1005. https://doi.org/10.1093/nar/18.4.999

LeBauer, D., Chamberlain, S., Foster, Z., Bartomeus, I., Black, C., Harris, D., Collins, R. (2024). traits: Species Trait Data from Around the Web. 10.32614/CRAN.package.traits

Lee, P. S., Sing, K. W., & Wilson, J. J. (2015). Reading mammal diversity from flies: the persistence period of amplifiable mammal mtDNA in blowfly guts (Chrysomya megacephala) and a new DNA mini-barcode target. *PloS one*, 10(4), e0123871. https://doi.org/10.1371/journal.pone.0123871

Lynggaard, C., Nielsen, M., Santos-Bay, L., Gastauer, M., Oliveira, G., & Bohmann, K. (2019). Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests. *Environmental DNA*, 1(4), 329-341. https://doi.org/10.1002/edn3.34

Marquina, D., Andersson, A. F., & Ronquist, F. (2018). New mitochondrial primers for metabarcoding of insects, designed and evaluated using in-silico methods. Molecular Ecology Resources, 19(1), 90-104. https://doi.org/10.1111/1755-0998.12942

Meusnier, I., Singer, G. A., Landry, J. F., Hickey, D. A., Hebert, P. D., & Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. BMC genomics, 9, 1-4. https://doi.org/10.1186/1471-2164-9-214

Mudalige, N. (2021). BOLD. R: A Software Package to Interface with BOLD Through R. In Recent Developments in Mathematical, Statistical and Computational Sciences: The V AMMCS International Conference, Waterloo, Canada, August 18–23, 2019 (pp. 187-196). Springer International Publishing. https://doi.org/10.1007/978-3-030-63591-6_18

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853-858. DOI: 10.1038/35002501

O'Donnell, J. L., Kelly, R. P., Lowell, N., & Port, J. A. (2016). Indexed pcr primers induce template-specific bias in large-scale dna sequencing studies. Plos One, 11(3), e0148698. https://doi.org/10.1371/journal.pone.0148698

Okonechnikov, K., Golosova, O., Fursov, M., & Ugene Team. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28(8), 1166-1167. https://doi.org/10.1093/bioinformatics/bts091

Oliveira, R. R., Silva, R., Nunes, G. L., & Oliveira, G. (2021). PIMBA: A pipeline for Meta Barcoding Analysis. In Advances in Bioinformatics and Computational Biology: 14th Brazilian Symposium on Bioinformatics, BSB 2021, Virtual Event, November 22–26, 2021, Proceedings 14 (pp. 106-116). Springer International Publishing. https://doi.org/10.1007/978-3-030-91814-9_10

Pacheco, M., Kajin, M., Gentile, R., Zangrandi, P. L., Vieira, M. V., & Cerqueira, R. (2013). A comparison of abundance estimators for small mammal populations. Zoologia (Curitiba), 30, 182-190. https://doi.org/10.1590/S1984-46702013000200008

Paglia, A. P., da Fonseca, G. A., Rylands, A. B., Herrmann, G., Aguiar, L., Chiarello, A. G., ... & Patton, J. L. (2012). Lista anotada dos mamíferos do Brasil. Occasional papers in conservation biology.

Piper, A. M., Batovska, J., Cogan, N. O., Weiss, J., Cunningham, J. P., Rodoni, B. C., & Blacket, M. J. (2019). Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience*, 8(8), giz092. https://doi.org/10.1093/gigascience/giz092

Pompanon, F., Deagle, B. E., Symondson, W. O., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: diet assessment using next generation sequencing. *Molecular ecology*, 21(8), 1931-1950. https://doi.org/10.1111/j.1365-294X.2011.05403.x

Primmer, C. R., Møller, A. P., & Ellegren, H. (1996). A wide-range survey of cross-species microsatellite amplification in birds. Molecular ecology, 5(3), 365-378. https://doi.org/10.1046/j.1365-294X.1996.00092.x

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system (www. barcodinglife. org). *Molecular Ecology Notes,* 7: 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Raven, P. H., Gereau, R. E., Phillipson, P. B., Chatelain, C., Jenkins, C. N., & Ulloa, C. U. (2020). The distribution of biodiversity richness in the tropics. Science Advances, 6(37). https://doi.org/10.1126/sciadv.abc6228

Reeves, L. E., Gillett-Kaufman, J. L., Kawahara, A. Y., & Kaufman, P. E. (2018). Barcoding blood meals: new vertebrate-specific primer sets for assigning taxonomic identities to host DNA from mosquito blood meals. PLoS neglected tropical diseases, 12(8), e0006767. https://doi.org/10.1371/journal.pntd.0006767

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic acids research, 39(21), e145-e145. https://doi.org/10.1093/nar/gkr732

Santos-Filho, M. D., Lázari, P. R. D., Sousa, C. P. F. D., & Canale, G. R. (2015). Trap efficiency evaluation for small mammals in the southern Amazon. Acta Amazonica, 45, 187-194. https://doi.org/10.1590/1809-4392201401953

Seitz, V., Schaper, S., Dröge, A., Lenze, D., Hummel, M., & Hennig, S. (2015). A new method to prevent carry-over contaminations in two-step pcr ngs library preparations. Nucleic Acids Research, gkv694. https://doi.org/10.1093/nar/gkv694

Sherrill-Mix, S. (2019). taxonomizr: Functions to Work with NCBI Accessions and Taxonomy. See https://CRAN. R-project. org/package= taxonomizr.

Stephenson, P. J. (2017). Small mammal monitoring: why we need more data on the Afrotheria. Oceania, 70(632), 11-1.

Stothard P (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28:1102-1104. https://doi.org/10.2144/00286ir01

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For biodiversity research and monitoring. Oxford University Press.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology*, 21(8), 2045-2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Taylor, P. G. (1996). Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. Molecular biology and evolution, 13(1), 283-285. https://doi.org/10.1093/oxfordjournals.molbev.a025566

Thermo Fisher Scientific Inc. (2024). Multiple Primer Analyzer. (Available at https://www.thermofisher.com/br/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools)

Thomsen, P. F., Kielgast, J. O. S., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., ... & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular ecology*, 21(11), 2565-2573. https://doi.org/10.1111/j.1365-294X.2011.05418.x

Townzen, J. S., Brower, A. V. Z., & Judd, D. D. (2008). Identification of mosquito bloodmeals using mitochondrial cytochrome oxidase subunit I and cytochrome b gene sequences. *Medical and veterinary entomology*, 22(4), 386-393. https://doi.org/10.1111/j.1365-2915.2008.00760.x

Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., ... & Miya, M. (2017). Environmental DNA enables detection of terrestrial mammals from forest pond water. Molecular Ecology Resources, 17(6), e63-e75. https://doi.org/10.1111/1755-0998.12690

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. Trends in ecology & evolution, 24(2), 110-117. https://doi.org/10.1016/j.tree.2008.09.011

Wei, X., Kuhn, D. N., & Narasimhan, G. (2003). Degenerate primer design via clustering. In Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003 (pp. 75-83). IEEE. doi: 10.1109/CSB.2003.1227306

Wilson, D. E., & Reeder, D. M. (Eds.). (2005). Mammal species of the world: a taxonomic and geographic reference (Vol. 1). JHU press. (Available at http://www.press.jhu.edu).

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden T (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 13:134. https://doi.org/10.1186/1471-2105-13-134

Statements and Declarations

This work was supported by CNPq, ICMBio and FAPEMIG (Grant CNPq/ICMBio/FAPs nº18/2017). B.R.N.C. received scholarship from Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). J.E.R.S. received a postdoctoral CAPES fellowship, and F.R.S. has a research fellowship from CNPq.

The authors have no relevant financial or non-financial interests to disclose.

All authors contributed to the study conception and design. FRS and JESJ have written the approved grant proposal of the metabarcoding project (CNPq 421303/2017-4). Material preparation, data collection and analyses were performed by BRNC. The first draft of the manuscript was written by BRNC and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability

The datasets generated during and/or analysed during the current study are available in the Github [https://github.com/babinecha/BrazilianMammalsDB], and NCBI/GenBank [Accession numbers from PQ529421 to PQ529445] or from the corresponding author on reasonable request.

Supplementary Information

Online Resource 1. Primer modifications, PCR conditions and recipes for Illumina Library construction and assembly, and Diagnostic tree detailed steps (**Fig 1**).

Online Resource 2. Spreadsheet containing the Brazilian mammal species list compiled from the list of Brazilian Society of Mastozoology (SBMz) (Abreu *et al.* 2023), complemented with Brazilian species according to the Mammal species of the world (Wilson & Reeder 2005), as well as exotic mammals present in Brazil (Da Rosa *et al.* 2017). It presents the taxonomic information of 830 mammal species from 270 genera, 55 families, and 11 orders, along with their name source and synonyms where available. The columns indicate whether the species is native or exotic to Brazil, its conservation status according to the IUCN Red List Version 2024-1 (IUCN 2024), its locomotion type (terrestrial "Te", aquatic "Aq", scansorial "Sc", semi-aquatic "SA", arboreal

"Ar", flying "Vo") according to Abreu *et al.* (2023), and occurrence across six Brazilian biomes: Amazon (AMAZ), Atlantic Forest (MATL), Cerrado (CERR), Caatinga (CAAT), Pantanal (PANT), Pampa (PAMP), and Marine (MARI), according to Abreu *et al.* (2023) and Paglia *et al.* (2012).

Online Resource 3. Spreadsheet containing the customized reference database for barcoding and metabarcoding Brazilian mammals. It includes sequence IDs obtained from GenBank and BOLD for COI, RNA12S, RNA16S, and mitogenomic sequences, along with the respective taxonomic information. The columns provide the following data: phylum, class, order, family, genus, species, source of name, accession number, sequence type, sequence source, and Brazilian mammal status (whether the species is native, exotic, or foreign to Brazil, or if the sequence belongs to other animal classes). It also presents the distribution of species across different Brazilian biomes, represented by the columns AMAZ (Amazon), MATL (Atlantic Forest), CERR (Cerrado), CAAT (Caatinga), PANT (Pantanal), PAMP (Pampa), and MARI (Marine). The values in the biome columns indicate the presence (1) or absence (0) of the species in each biome.

Online Resource 4. Table containing all compiled and modified primers analysed in this study, along with their respective sequences and characteristics for mitochondrial markers used in metabarcoding of mammal species. It specifies the amplifiable genetic marker for each primer, including the primer type (forward or reverse), name, reference, nucleotide sequences (5'-3'), and lengths. The chemical characteristics of the primers, including the minimum and maximum percentage of guanine-cytosine content (%GC) and minimum and maximum melting temperatures (TM), were obtained using Primer Stats (Stothard, 2000).

Online Resource 5. Spreadsheet containing results of *in-silico* PCR performed using the *primerTree* package in R (Cannon *et al.* 2016). It details the number of sequences retrieved and taxonomically identified from GenBank using each primer pair listed in Online Resource 4. The data is organized by genetic marker, primer pair (forward and reverse), taxonomic classification (class, order, family), number of species, number of sequences, product lengths, and the number of mismatches between the primer and target sequences.

Online Resource 6. Spreadsheet containing results from *in-vitro* PCR performed with three genetic markers: 1) 205 bp of the COI gene (modified primers from Lee *et al.* 2015), 2) 171 bp of RNA12S (original primers from Ushio *et al.* 2017), and 3) 78 bp of RNA16S (modified primers from Haile *et al.* 2009). Two amplification techniques were applied to mammalian samples: Sanger sequencing and Illumina sequencing (using a two-step PCR protocol adapted from Ushio *et al.* 2017 and Chen *et al.* 2021, followed by the PIMBA pipeline from Oliveira *et al.* 2021). The spreadsheet includes Sample ID, DNA mix ID, collection information, provisional taxonomic classifications based on morphology (order, family, genus, species), applicable synonyms, number of sequence present in the customized Brazilian mammalian database (for COI, 12S, and 16S), and sequencing results. For each sample, species identification, percent identity, number of reads (Illumina only), reference database used, and diagnostic information are provided.

3. Capítulo II. Análise da biodiversidade de vertebrados detectada através de *metabarcoding* de iDNA em biomas brasileiros (artigo não submetido)

RESUMO

Os biomas neotropicais, incluindo a Mata Atlântica e o Cerrado, são hotspots de biodiversidade que enfrentam ameaças severas, indicando a necessidade de métodos eficazes de monitoramento. O *metabarcoding* de iDNA oferece uma alternativa poderosa aos levantamentos tradicionais, analisando DNA extraído de invertebrados. Aqui, aplicamos essa metodologia pela primeira vez na Mata Atlântica e no Cerrado, além de locais na Amazônia, utilizando mutucas (Tabanidae) e moscas varejeiras (Calliphoridae e Sarcophagidae) como amostradores. As identificações taxonômicas basearam-se em três marcadores genéticos (COI, RNA12S, RNA16S) e um banco de referência, ambos customizados para mamíferos brasileiros. Identificamos 106 unidades taxonômicas operacionais (OTUs) abrangendo mamíferos, aves, anfíbios, répteis e peixes, incluindo espécies elusivas, uma ameaçada (*Sapajus nigritus*), uma exótica (*Lepus sp.*) e uma doméstica (*Bos taurus*). O RNA16S foi o marcador que mais contribuiu para a diversidade taxonômica, enquanto o COI e o RNA12S foram mais eficazes para mamíferos e aves. A eficiência de amostragem variou com a sazonalidade, e o agrupamento de moscas detectou uma gama mais ampla de táxons, embora moscas individuais também tenham contribuído para a biodiversidade registrada. Os desafios incluíram bancos de referência incompletos, contaminação cruzada e eficiência reduzida de amplificação. Este estudo destaca o potencial do *metabarcoding* de iDNA para complementar levantamentos tradicionais, oferecendo insights escaláveis sobre a biodiversidade de vertebrados em biomas neotropicais. Ao expandir bancos de referência e melhorar metodologias, essa abordagem pode guiar estratégias de conservação e apoiar avaliações de saúde ecossistêmica em *hotspots* de biodiversidade.

Title
iDNA Metabarcoding for Biodiversity Monitoring in Neotropics: A Case Study in Brazilian Biomes

Author information
Barbara R N Chaves (barbarachaves@ufmg.br, https://orcid.org/0009-0008-0977-8974)
Jose Eustaquio Santos-Junior (jrsantos140782@yahoo.com.br, https://orcid.org/0000-0002-7150-3751)
Fabricio R Santos (fsantos@icb.ufmg.br, https://orcid.org/0000-0001-9088-1750)
Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brazil

ABSTRACT

Neotropical biomes, including the Atlantic Forest and Cerrado, are biodiversity hotspots facing severe threats, highlighting the need for effective monitoring. iDNA metabarcoding offers a powerful alternative to traditional surveys by analyzing DNA from invertebrates. Here, we applied this method for the first time in the Atlantic Forest and Cerrado, along with Amazonian sites, using horseflies (Tabanidae) and carrion flies (Calliphoridae and Sarcophagidae) as samplers. Taxonomic assignments relied on three genetic markers (COI, RNA12S, RNA16S) and a customized reference database for Brazilian mammals. We identified 106 operational taxonomic units (OTUs) spanning mammals, birds, amphibians, reptiles, and fish, including a threatened (*Sapajus nigritus*), an exotic (*Lepus sp.*), and a domestic (*Bos taurus*) species. RNA16S contributed the most to taxonomic diversity, while COI and RNA12S were effective for mammals and birds. Sampling efficiency varied with seasonality, and pooling flies detected a broader range of taxa, though individual flies also added biodiversity. Challenges included incomplete reference databases, cross-contamination, and reduced amplification efficiency. This study highlights the potential of iDNA metabarcoding to complement traditional surveys, offering scalable insights into vertebrate biodiversity in Neotropical biomes. By expanding reference databases and improving methodologies, this approach can guide conservation strategies and support ecosystem health assessments in biodiversity hotspots.

INTRODUCTION

Neotropical biomes, such as the Amazon, Atlantic Forest, and Cerrado, are among the most degraded and rapidly disappearing ecosystems globally. Recognized as biodiversity hotspots, they harbor exceptional species richness while facing severe threats (Myers et al 2000; Jenkins et al 2013). Despite their importance, these ecosystems are disproportionately under-sampled (Hughes et al 2021), leading to numerous Neotropical species undergoing local extinctions or vanishing before being formally described. Although their relatively large body size compared to other groups, mammals are challenging to study in the Neotropical region due to limited accessibility and logistical constraints, as well as low population densities and the elusive nature of many species due to their ability to conceal themselves within the complex tropical vegetation (Schipper et al 2008; Schnell et al 2012). Although mammals receive significant attention in conservation research, 13% of mammal species are still classified as Data Deficient (DD) by the IUCN, with 35% occurring in the Neotropics (IUCN, 2024).

This highlights the urgent need to expand taxonomic inventories and knowledge of natural history (Thomsen et al 2012). Traditional taxonomic approaches, however, demand extensive fieldwork and require specialized expertise in morphology and taxonomy. Metabarcoding of invertebrate-derived DNA (iDNA) has emerged as a promising tool for biodiversity surveys, enabling the detection of elusive and rare vertebrate species through the analysis of DNA extracted from invertebrates that feed on vertebrates (Calvignac-Spencer et al., 2013). Metabarcoding of iDNA presents similar or higher efficiency in detecting vertebrates than conventional methods (e.g. camera trapping) and can generate species lists that support ecological analyses and biodiversity monitoring (Keck et al., 2023). Unlike mammals, invertebrates can be easily collected, for instance, through flight interception traps (e.g. Malaise traps), requiring minimal sampling effort to capture large numbers of diverse fly species (Blahó et al 2013; Lynggaard et al 2019; Skvarla et al 2021).

Although most iDNA metabarcoding studies have been conducted in temperate zones, reflecting limited research funding in many tropical countries (Carvalho et al 2022), it has been successfully applied in the Neotropics, within a growing number of studies utilizing diverse invertebrate groups as sources of vertebrate DNA (e.g., Kocher et al 2017; Rodgers et al 2017; Lynggaard et al 2019; Massey et al 2022; Saranholi et al 2024). However, the Atlantic Forest, for instance, despite its importance for conservation—second richest biome in terms of mammalian diversity, harboring about 300 mammal species, 30% of which are endemic (Paglia et al, 2012), and facing severe anthropogenic pressures such as habitat loss and defaunation (Bogoni et al., 2018; Galetti et al., 2017)—remains understudied, with no published iDNA surveys to date, with exception of a zoo in a transition area between Cerrado and Atlantic Forest (Saranholi et al., 2023).

Each iDNA source can introduce taxonomic biases, due to differences in feeding ecology and life history (Massey et al 2022). However, flies from Calliphoridae and Sarcophagidae families (carrion flies) feed on dead animals, open wounds, and feces, and have been frequently used as iDNA samplers, presenting no noticeable taxonomic bias: they enable the detection of a wide diversity of vertebrate species, including terrestrial, volant, and arboreal mammals (Calvignac-Spencer et al 2013; Rodgers et al 2017; Gogarten et al 2020; Lee et al 2023). Hematophagous flies from the family Tabanidae (horseflies) also have potential as iDNA sources, as they are highly adapted for widespread dispersion, using their flight ability to locate hosts and reach even the most elusive mammals of tropical environments (Brown, 2020). A recent work successfully identified mammal and bird species using mosquitoes and flies from several families (Saranholi et al. 2023), but horseflies have not yet been used as vertebrate samplers.

The selection of appropriate genetic markers is crucial for metabarcoding efficiency, particularly in hyperdiverse regions like the Neotropics, where high genetic variability can hinder amplification and taxonomic identification (Carvalho et al., 2022). While ribosomal genes (e.g., RNA12S and RNA16S) are commonly used, they can lack species-level resolution for some groups and are supported by limited reference databases (Alberdi et al., 2018). The COI gene, on the other hand, offers higher taxonomic resolution and a broader database but requires careful primer design to ensure compatibility with local biodiversity using short sequences (Hajibabaei, 2012; Clarke et al., 2017). Combining multiple markers can mitigate individual limitations, improving both detection and taxonomic resolution (Teixeira et al 2023).

In this study, we aimed to survey mammal diversity through iDNA across three Neotropical biomes (Amazon, Atlantic Forest, and Cerrado), maximizing detection by sampling flies from three families (Calliphoridae, Sarcophagidae, and Tabanidae). We employed a combination of primers targeting three genetic markers (COI, RNA12S, and RNA16S), alongside a customized reference database of known Brazilian mammal species, which demonstrated potential for species-level identification based on *in silico* and *in vitro* analyses (Chaves et al., submitted). This study not only serves as a pilot to validate the performance of these primers in real-world conditions but also represents a critical step toward expanding the application of metabarcoding in biodiversity monitoring and conservation in the Neotropics.

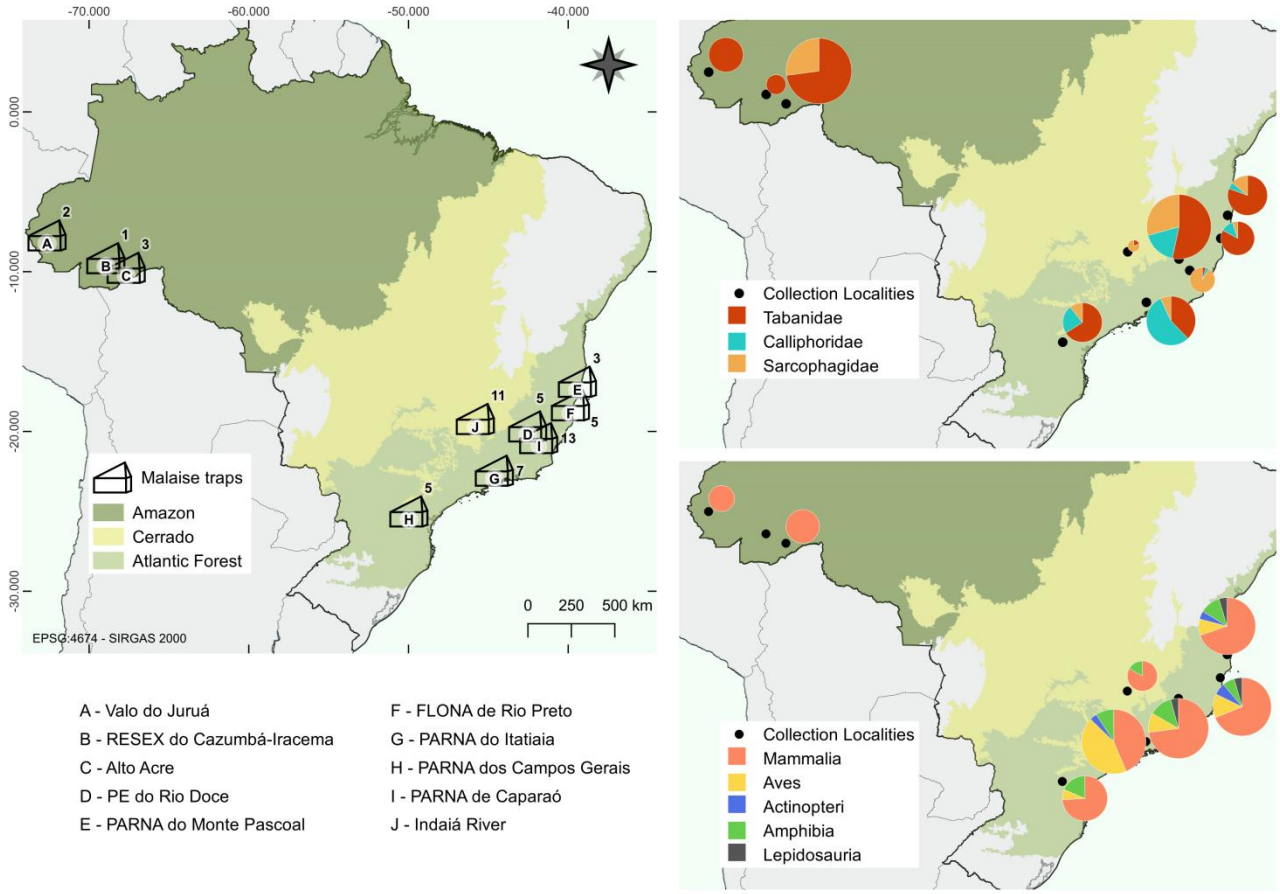

Fig 1 Maps of sampled localities of iDNA metabarcoding of horseflies and carrion-flies, sampled across three Neotropical biomes (Amazon, Atlantic Forest, and Cerrado): **a)** number of Malaise traps/points per sampled

locality; **b)** proportion of sampled flies per locality and fly families (Tabanidae, Sarcophagidae, and Calliphoridae); **c)** proportion of vertebrate OTUs detected by locality and vertebrate classes (Mammalia, Aves, Actinopteri, Amphibia, and Lepidosauria).

METHODS

Survey localities and Insect sampling

Fieldwork was conducted by a team of nine people across six points in three localities in the Amazon biome, all in Acre State (Juruá Valley, RESEX Cazumbá-Iracema, and Alto Acre), six points in the Cerrado biome, all in Minas Gerais State (region of Indaiá River), and 29 points in the Atlantic Forest biome, across five Brazilian states and six conservation units (PARNA Monte Pascoal, FLONA Rio Preto, PE Rio Doce, PARNA Caparaó, PARNA Itatiaia, and PARNA Campos Gerais). Conservation units were used as collection localities due to their crucial role in preserving biodiversity by helping maintain ecological processes and protect species from anthropogenic pressures. In total, five expeditions were conducted across 41 sampling points within 10 localities (Fig 1a), resulting in 369 trap days of sampling and encompassing diverse environmental conditions and geographic regions.

Insects were collected passively using unbaited Malaise-type flight interception traps with absolute ethanol-filled bottles. Traps were placed at each collection locality and left for an average of five days (minimum one day and maximum 13 days) during five sampling periods: July 2017, June and September 2018, November 2019, and December 2022. Variations in trap exposure and sampling periods occurred due to logistical challenges, but environmental conditions during different seasons are known to influence fly abundance and diversity, potentially affecting sampling efficiency. At PE Rio Doce, black balloons were positioned near the traps to enhance the collection of horseflies, which are visually attracted to large, dark objects (Brown, 2020; Skvarla et al., 2021). All collected samples were preserved in absolute ethanol at -20°C to ensure the integrity of DNA for subsequent analyses, and stored in the Laboratory of Biodiversity and Molecular Evolution of UFMG (LBEM-UFMG).

Fly specimens from families Calliphoridae, Sarcophagidae, and Tabanidae were sorted and individualized from each sample under stereomicroscopes at the Laboratory of Insect Systematics of UFMG (LSI-UFMG). The intestines of larger specimens and the entire abdomens of smaller flies considered the most likely sources of vertebrate iDNA due to their contact with ingested material, were carefully dissected using sterilized razor blades and tweezers. The tissues were individually digested in a lysis mix (Wilson, 2012) and subjected to DNA extraction using DNeasy Blood & Tissue Kits (QIAGEN), following the manufacturer's protocols.

The dissections and DNA extractions were performed in a dedicated extraction room at LBEM-UFMG, with all surfaces and equipment thoroughly cleaned with 10% bleach and RNase AWAY before and after each session to prevent contamination. A negative control, consisting of reagents without tissue samples, was included for every 23 samples to monitor contamination. Extracted DNA samples were quantified and assessed for purity using a NanoDrop spectrophotometer (Thermo Scientific), with A260/280 ratios recorded to ensure sample quality.

Metabarcoding of iDNA

Metabarcoding of iDNA samples followed a two-step PCR protocol for Illumina library preparation, adapted from Ushio et al. (2017) and Chen et al. (2021), as detailed in Chaves et al. (submitted). Three genetic markers were amplified using primers designed or specifically adapted for mammals: i) 205 bp of COI (Chaves et al., submitted; modified from Lee et al., 2015), ii) 171 bp of RNA12S (Ushio et al., 2017), and iii) 78 bp of RNA16S (Chaves et al., submitted; modified from Haile et al., 2009). All PCRs were prepared in a dedicated space at LBEM-UFMG, where all surfaces and equipment were previously UV-sterilized and decontaminated with 10% bleach and RNase AWAY. Each reaction was performed in duplicate, with a negative control included for every 23 samples to monitor contamination. Positive controls, consisting of DNA mixes from 16 mammals (Table 1), were included to verify reaction efficacy and to evaluate the quality and intensity of amplified products. The reaction mixes and cycling programs are described in Chaves et al. (submitted)

Table 1. List of mammals used as positive controls for metabarcoding of iDNA.

Order	**Family**	**Species**
Carnivora	Felidae	*Puma yagouaroundi*
	Mustelidae	*Lontra longicaudis*
Cetacea	Delphinidae	*Sotalia fluviatilis*
Chiroptera	Emballonuridae	*Rhynchonycteris naso*
	Molossidae	*Molossops temminckii*
	Noctilionidae	*Noctilio albiventris*

	Phyllostomidae	*Artibeus lituratus*
		Desmodus rotundus
Didelphimorphia	Didelphidae	*Didelphis albiventris*
Pilosa	Bradypodidae	*Bradypus torquatus*
	Myrmecophagidae	*Myrmecophaga tridactyla*
		Tamandua tetradactyla
Primates	Atelidae	*Alouatta guariba*
Rodentia	Cricetidae	*Oligoryzomys fornesi*
	Cuniculidae	*Cuniculus paca*
Sirenia	Trichechidae	*Trichechus manatus*

To reduce costs, DNA from collected and sorted flies were pooled into template mixes before amplification. However, as pooling may reduce the number of species detected (Rodgers et al., 2017), specimens from visually amplified template mixes were also individually amplified for the same genetic markers. Unique 7-base tag identifiers (Hamady et al., 2008) were added to both ends of the PCR amplicons, including positive controls, enabling posterior identification of samples. Properly labeled products were normalized, pooled, and sequenced on the Illumina MiSeq platform at the Genomics Laboratory (LG-UFMG) using the MiSeq Reagent Kit v2 (Illumina), with 25% PhiX added for quality control.

Illumina output was processed using the PIMBA pipeline (Oliveira et al., 2021), which performs demultiplexing, quality filtering, Operational Taxonomic Unit (OTU) clustering, error correction, and taxonomic assignment of reads. Taxonomic assignments were based on similarity, using a newly customized reference database (Chaves et al., submitted), which encompasses a combination of COI, RNA12S, and RNA16S sequences from 72% of known Brazilian mammal species and 87% of genera across all orders and biomes, as well as extant exotic species in Brazil. OTUs without identification, unexpected matches, or sequence identities below 95% were reanalyzed via BLASTn (Altschul et al., 1990) against the full GenBank database, retaining only the best matches. OTUs shorter than 100 bp, longer than 260 bp, or deviating by more than 19 bp from the expected marker size were excluded, along with OTUs represented by unique reads.

To refine taxonomic assignments, OTUs processed by PIMBA were further analyzed using the PROTAX pipeline (Somervuo et al., 2016), with the same reference database. PROTAX accounts for the incompleteness of taxonomic and reference databases and calculates probabilities for taxonomic classification, improving the accuracy of assignments initially based on sequence similarity. The final taxonomy of each OTU was determined by combining sequence similarity and PROTAX-derived probabilities thresholds. Assignments were considered valid when both PIMBA and PROTAX pointed out to the same taxon, when probabilities exceeded a threshold (> -1) at a given taxonomic level and sequence similarities met the following criteria: $\geq 98\%$ for species level, $\geq 95\%$ for genus level, $\geq 90\%$ for family level, and $\geq 80\%$ for order level (Fig 2). Additionally, information on species distribution (Chaves et al, submitted for mammals, and Piacentini et al 2015 for birds) was used to validate or correct assignments, raising the taxonomic levels when inferred taxa were not expected to occur in Brazil or in the sampled biome, ensuring alignment with known biogeographical data. In the absence of information on species occurrence in Brazil, OTUs were assigned to the family level or higher. Similarly, in the absence of biome-specific occurrence data, OTUs were assigned to the genus level or higher.

High-throughput sequencing (HTS), such as Illumina sequencing, is highly susceptible to contamination events, including the release of aerosolized amplicons and sample carry-over during the manipulation of multi-round PCR products. These issues can lead to tag jumping, where the pooling of multiple tagged amplicons before sequencing can lead to unintended recombination of tags and result in labeling sequences to incorrect samples (Murray et al., 2015). Unlike Sanger sequencing, where contaminants can obscure real sequences, HTS detects contaminants as unambiguous sequences, potentially leading to false positives. To minimize false positive assignments arising from cross-contamination, particularly from positive controls, OTUs detected in iDNA samples were systematically compared to control species. Cross-contamination was assessed and resolved according to a defined workflow (Fig 3).

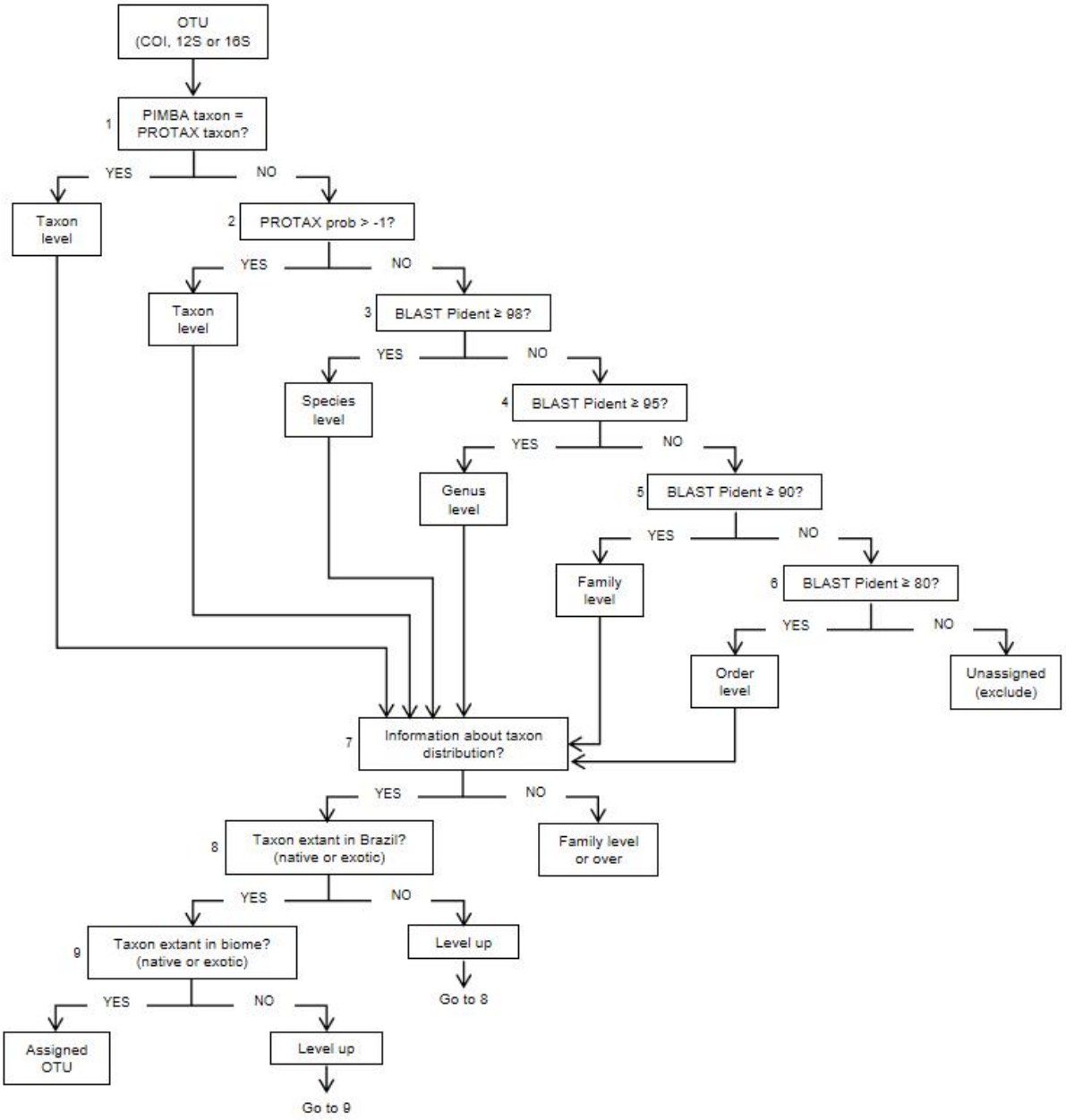

Fig 2 Molecular taxonomic key presenting the steps for correct assignment of metabarcoding sequences (COI, RNA12S or RNA16S), considering BLAST similarities (Pident) obtained with PIMBA pipeline (Oliveira et al., 2021), taxon probabilities obtained with PROTAX pipeline (Somervuo et al., 2016) and information about species distribution (extant in Brazil and extant in biome).

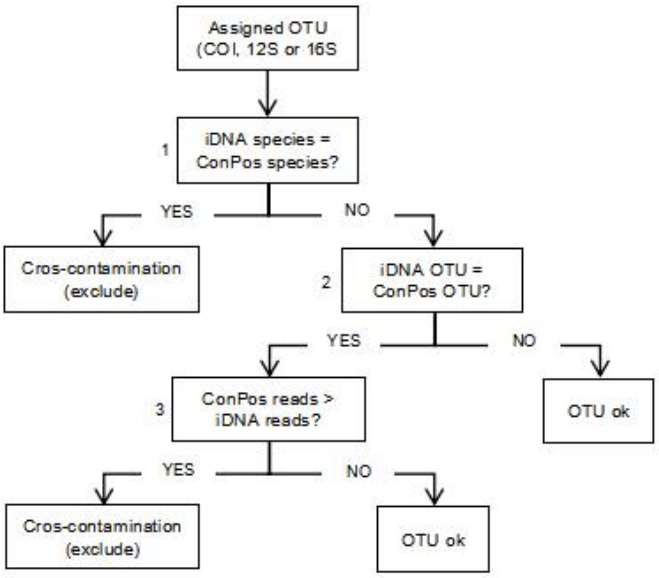

Fig 3 Dichotomous key presenting the steps for identification of cross-contamination of assigned metabarcoding sequences (COI, RNA12S, or RNA16S), considering correspondence among iDNA and control species or OTUs.

RESULTS

Across the 41 sampling points spanning diverse environmental conditions and geographic regions, 631 flies were collected and sorted, averaging 63 flies per locality or 1.7 flies per trap day. These included 402 Tabanidae, 96 Calliphoridae, and 133 Sarcophagidae (Fig 1b). The number of flies collected at each locality reflected variations in sampling effort; however, specific environmental characteristics also appeared to influence the results. For instance, despite low sampling effort, large numbers of flies were collected in Amazonian localities (approximately 30 flies per trap-day), whereas high sampling efforts at Indaiá River (Cerrado) and PARNA Itatiaia (Atlantic Forest) yielded relatively few samples (0.1 and 0.6 flies per trap-day, respectively). The least localities were collected during the winter, when usually dry and low temperatures may have reduced fly availability. The use of black balloons as attractants at PE Rio Doce (Atlantic Forest) appeared to enhance horsefly captures, as this locality recorded the highest number of these flies (80), except for Alto Acre in the Amazon, which had the most (113).

DNA extracted from flies was pooled into 51 template mixes, 26 of which were successfully amplified and sequenced, along with 106 individual flies, representing approximately 17% of the total flies (11% of Tabanidae, 25% of Calliphoridae, and 9% of Sarcophagidae). After three runs, Illumina sequencing generated a total of 19,904,256 sequence reads, reduced to 2,087,007 reads and 280 OTUs after demultiplexing, quality filtering, clustering, error correction, count filtering, and length filtering (Table 2). Most of the reads originated from positive controls (56%), leaving 921,188 reads from iDNA samples and template mixes. After comparing OTUs from iDNA and control samples, 30% of total iDNA reads were identified as cross-contamination and excluded, along with 1,743 unassigned reads. This resulted in 643,947 reads and 155 OTUs, including 22 from COI, 41 from RNA12S, and 92 from RNA16S (Table 2), with RNA12S contributing the highest read counts. The majority of OTUs (73%) were identified as vertebrates, although some were microorganisms or invertebrates, likely representing species concurrently sampled with flies in the Malaise traps. Notably, five OTUs (from RNA12S and RNA16S) were assigned to the dipteran families Sarcophagidae or Tabanidae, likely due to the amplification of fly specimens from which the iDNA was extracted.

Table 2. Summary of sequencing counting generated by Illumina for each genetic marker, including total numbers of corrected reads, reads from positive controls, unassigned reads, reads identified as cross-contamination, valid reads (after filtering and corrections), besides numbers of valid OTUs, and the average number of reads per OTU.

Marker	Corrected reads	Unassigned reads	ConPos reads	Cross-Contam reads	Valid reads	Valid OTUs	reads/ OTUs
COI	400,783	137	287,020	80,973	32,878	22	1,494
RNA12S	766,811	309	151,009	105,925	509,985	41	12,439
RNA16S	919,413	1,317	727,790	90,370	101,084	92	1,099
Total	2,087,007	1,763	1,165,819	277,268	643,947	155	4,154

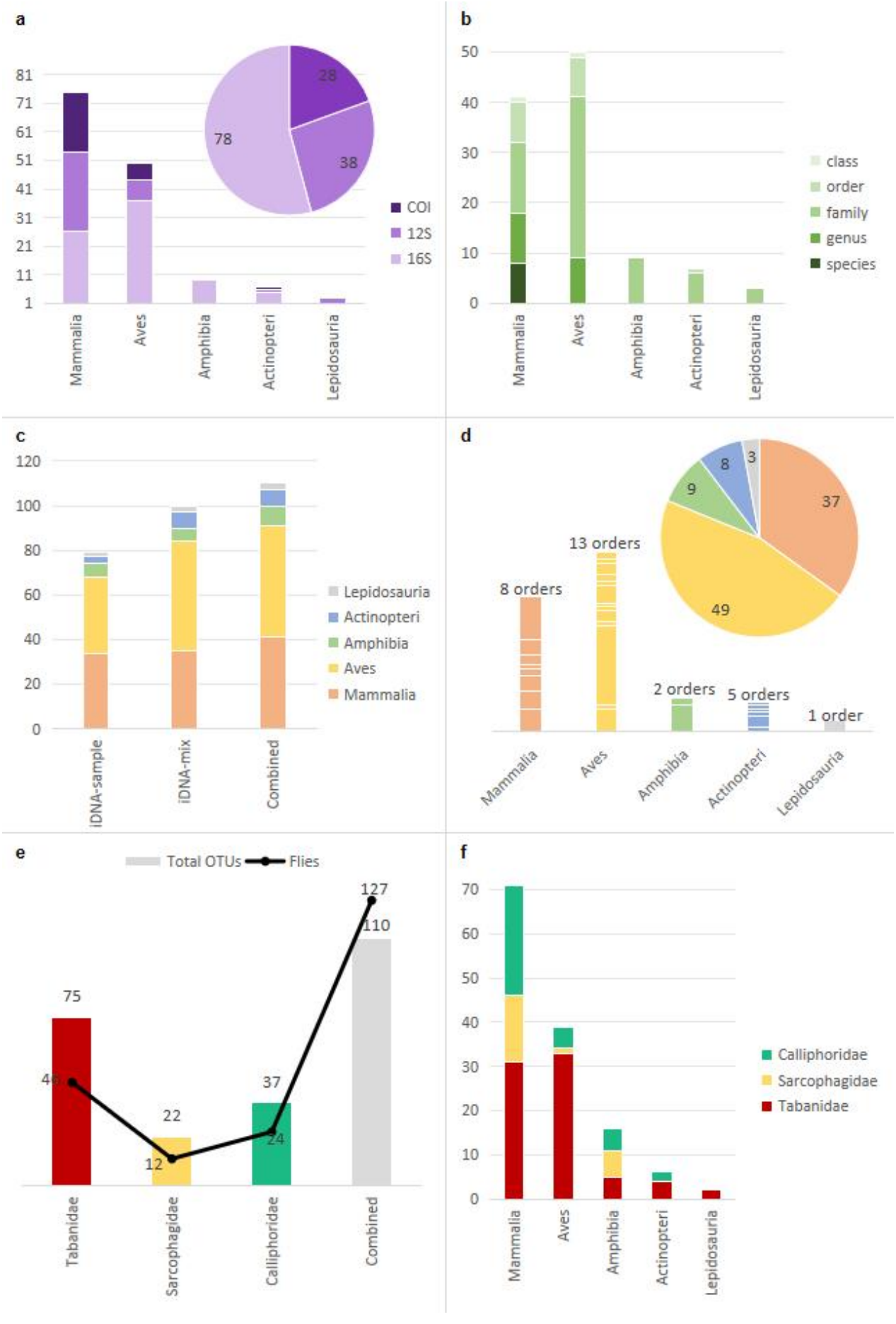
a
81
71
61
51
41
31
21
11
1
28
38
78
COI
12S
16S
Mammalia
Aves
Amphibia
Actinopteri
Lepidosauria
b
50
40
30
20
10
0
class
order
family
genus
species
Mammalia
Aves
Amphibia
Actinopteri
Lepidosauria
c
120
100
80
60
40
20
0
Lepidosauria
Actinopteri
Amphibia
Aves
Mammalia
iDNA-sample
iDNA-mix
Combined
d
3
8
9
37
49
13 orders
8 orders
2 orders
5 orders
1 order
Mammalia
Aves
Amphibia
Actinopteri
Lepidosauria
e
Total OTUs
Flies
127
110
75
46
37
22
12
24
Tabanidae
Sarcophagidae
Calliphoridae
Combined
f
70
60
50
40
30
20
10
0
Calliphoridae
Sarcophagidae
Tabanidae
Mammalia
Aves
Amphibia
Actinopteri
Lepidosauria

Fig 4 Characterization of OTUs obtained from iDNA metabarcoding of horseflies and carrion-flies, sampled across three Neotropical biomes (Amazon, Atlantic Forest, and Cerrado): **a)** OTUs per genetic marker (COI, RNA12S, and RNA16S) and vertebrate classes (Mammalia, Aves, Amphibia, Actinopteri, and Lepidosauria); **b)** OTUs per assigned taxonomic level (species, genus, family, order, and class) and per vertebrate classes; **c)** OTUs obtained by sequencing individual samples (iDNA-sample), by pooled samples (iDNA-mix) and combined; **d)** OTUs and number of orders per vertebrate classes; **e)** number of sequenced flies and OTUs per fly families (Tabanidae, Sarcophagidae, and Calliphoridae), and combined; **f)** OTUs per fly families and vertebrate classes.

Slightly different quantities and compositions of vertebrate OTUs were obtained from individual fly samples and pooled iDNA (Fig 4c). While individual samples detected 79 vertebrate taxa, pooled samples identified 100. However, some OTUs were exclusive to either individual or pooled samples, demonstrating that combining both approaches increased the overall diversity detected. Among the vertebrate OTUs, six were assigned to humans, representing 7% of valid reads. While these may originate from iDNA, as flies are known to feed on humans, contamination during sample handling in the field or laboratory is also a significant possibility.

Although this study focused on mammal detection, vertebrates from other classes were also identified in the sampled iDNA, representing 65% of the vertebrate OTUs (Fig 4d). Table 3 provides a complete list of the vertebrate taxa detected. Unexpectedly, birds were the most frequently detected, with 49 OTUs (excluding humans), followed by mammals (41 OTUs), amphibians (9 OTUs), fishes (8 OTUs), and reptiles (3 OTUs). The diversity of vertebrates detected through iDNA is further illustrated by the large number of distinct orders identified (Fig 4d and Fig 1c), with birds being the most diverse, encompassing 13 orders, followed by mammals (8 orders), fishes (5 orders), amphibians (2 orders), and reptiles (1 order). Approximately three-quarters of OTUs from COI and RNA12S were assigned to mammals, with some assigned to birds, while OTUs from RNA16S, besides being more abundant, also captured a greater diversity of birds, amphibians, and fishes (Fig 4a). No single marker identified all vertebrates detected; however, their combined use resulted in a larger diversity of taxa.

The number of OTUs obtained per dipteran family reflected the abundance of flies successfully sequenced, with 75 vertebrate OTUs for Tabanidae, 22 for Sarcophagidae, and 37 for Calliphoridae, averaging 1.7 vertebrates per fly. The composition of vertebrates detected by each dipteran family varied, with the majority of mammals and birds identified through horseflies (Tabanidae), which also accounted for most of the observed vertebrate diversity. However, certain OTUs were exclusively detected by carrion flies (Sarcophagidae and Calliphoridae), underscoring that sampling multiple fly families enhances vertebrate detection (Fig 4e and Fig 4f).

In the absence of information on species presence in Brazil, all amphibians, reptiles, and fishes were assigned to the family level or higher. Similarly, in the absence of distribution data across Brazilian biomes, all birds were assigned to the genus level or higher. Consequently, most OTUs were assigned at the family level or higher (75%) (Fig. 4b), and, excluding humans, only eight OTUs were identified at the species level: three as *Bos taurus*, one as *Pecari tajacu*, one as *Potos flavus*, one as *Pteronotus gymnonotus,* and two as *Sapajus nigritus*. The detected mammals exhibited diverse habits, including terrestrial, arboreal, volant, scansorial, and fossorial species. Rodents were the most frequently detected group, with 12 OTUs, followed by artiodactyls (6 OTUs), carnivores (5 OTUs), bats and primates (4 OTUs each), pilosans (3 OTUs), American marsupials (Didelphimorphia, 2 OTUs), and an exotic rabbit (1 OTU).

The domestic cow (*Bos taurus*), detected by all three fly families in the Atlantic Forest and Amazon, was amplified by all markers and accounted for 19% of all valid reads. Also detected by three fly families in the Atlantic Forest, the black-horned capuchin (*Sapajus nigritus*) has a relatively restricted distribution and is currently listed as Near Threatened by the IUCN (2024). The other species identified are listed as Least Concern and have broad distributions across the Neotropics; while the collared peccary (*Pecari tajacu*) is terrestrial and lives in large social groups, the kinkajou (*Potos flavus*) is arboreal, with nocturnal and mostly solitary habits.

Table 3. List of vertebrate taxa identified in iDNA samples, including taxonomic classification (Class, Order, Family, Genus, Species), number of OTUs assigned to each taxon, genetic markers that contributed to the detection, and the source of iDNA based on the dipteran families used for sampling (Tabanidae, Sarcophagidae, Calliphoridae). Taban: Tabanidae; Sarco: Sarcophagidae; Calli: Calliphoridae.

Class	Order	Family	Genus	Species	n.OTUs	Markers	iDNA source
Actinopteri	Gadiformes	*unknow*			1	16S	*unknow*
		Gobiidae			2	16S	Taban, Calli
		Odontobutidae			1	16S	Taban
	Perciformes	Sciaenidae			1	16S	*unknow*

	Siluriformes	Loricariidae			1	12S	Taban
	Uranoscopiformes	Ammodytidae			1	COI	*unknow*
Amphibia	Anura	Bufonidae			3	16S	Taban, Sarco, Calli
		Hylidae			2	16S	Taban, Sarco, Calli
		Leptodactylidae			1	16S	*unknow*
		Microhylidae			1	16S	*unknow*
	Caudata	Ambystomatidae			1	16S	Sarco
		Plethodontidae			1	16S	*unknow*
Aves	*unknow*				1	16S	*unknow*
	Accipitriformes	*unknow*			3	12S, 16S	Taban
	Accipitriformes	Accipitridae	*unknow*		2	16S	*unknow*
	Accipitriformes	Accipitridae	*Milvus*		1	16S	*unknow*
	Anseriformes	Anatidae			1	16S	*unknow*
	Apodiformes	Trochilidae			22	COI, 12S, 16S	Taban
	Caprimulgiformes	*unknow*			1	COI	*unknow*
	Falconiformes	*unknow*			1	16S	*unknow*
	Falconiformes	Falconidae	*Falco*		2	16S	Taban, Calli
	Galliformes	*unknow*			1	12S	*unknow*
	Gruiformes	Rallidae	*Fulica*		1	16S	Taban
	Passeriformes	Corvidae			1	16S	*unknow*
		Pipridae	*unknow*		1	16S	*unknow*
		Pipridae	*Chiroxiphia*		1	COI	Calli
		Pipridae	*Manacus*		1	16S	*unknow*
		Tyrannidae	*Sublegatus*		1	12S	Taban
	Procellariiformes	Procellariidae			1	16S	*unknow*
	Psittaciformes	Psittacidae			2	COI, 12S	Taban
	Strigiformes	Strigidae	*unknow*		1	16S	Taban
			Athene		1	16S	*unknow*
			Bubo		1	16S	*unknow*
	Suliformes	*unknow*			1	16S	*unknow*
	Tinamiformes	*unknow*			1	COI	Calli
	Tinamiformes	Tinamidae			1	COI	Taban, Sarco, Calli
Lepidosauria	Squamata	Agamidae			2	12S, 16S	Taban
		Hydrophiidae			1	12S	*unknow*
Mammalia	*unknow*				1	16S	*unknow*
	Artiodactyla	*unknow*			1	16S	Taban
		Bovidae	*Bos*	*unknow*	1	COI	Taban, Sarco, Calli
				Bos taurus	3	COI, 12S, 16S	Taban, Sarco, Calli
		Tayassuidae	*Tayassu*	*Pecari tajacu*	1	12S	Taban, Sarco, Calli
	Carnivora	Canidae			1	16S	*unknow*
		Mustelidae			1	16S	Taban
		Procyonidae	*Nasua*		1	12S	*unknow*
			Potos	*Potos flavus*	1	16S	*unknow*
			Procyon		1	COI	Calli
	Chiroptera	*unknow*			1	COI	Taban, Calli
		Molossidae			1	16S	Taban, Sarco, Calli
		Mormoopidae	*Pteronotus*	*unknow*	1	16S	Calli
				Pteronotus gymnonotus	1	12S	Taban, Calli
	Didelphimorphia	*unknow*			1	12S	Taban
		Didelphidae	*Didelphis*		1	16S	*unknow*
	Lagomorpha	Leporidae	*Lepus*		1	16S	Taban, Sarco, Calli
	Pilosa	*unknow*			2	COI, 12S	Taban, Calli
		Bradypodidae	*Bradypus*		1	12S	Sarco, Calli

	Myrmecophagidae	*Myrmecophaga*		1	16S	Taban, Calli
Primates	*unknow*			2	COI, 12S	Taban, Calli
	Cebidae	*unknow*		2	12S, 16S	Taban, Sarco, Calli
		Sapajus	*Sapajus nigritus*	2	12S	Taban, Sarco, Calli
	Hominidae	*Homo*	*Homo sapiens*	6	COI, 12S	Taban, Sarco, Calli
Rodentia	*unknow*			1	COI	Taban, Calli
	Caviidae			1	12S	Taban, Sarco, Calli
	Cricetidae			6	COI, 12S, 16S	Taban
	Ctenomyidae	*Ctenomys*		1	COI	*unknow*
	Dasyproctidae	*Myoprocta*		1	12S	*unknow*
	Echimyidae			2	12S	Taban, Sarco, Calli

DISCUSSION

The iDNA metabarcoding approach used in this study enabled the detection of a wide range of mammal species, from conspicuous mammals—characterized by larger body sizes and/or social behavior (e.g., domestic cow, collared peccary, coatis)—to elusive mammals, such as volant, arboreal, scansorial, fossorial, and small-bodied species. Notably, this included a threatened species (*Sapajus nigritus*) and an exotic taxon (*Lepus sp*.). Detecting both endangered and invasive species is critical for informing conservation policies, as it supports prioritizing conservation actions, identifying habitats for protection or restoration, and managing invasive species, aligning with global biodiversity frameworks (Meyerson et al., 2022; Arponen, 2012).

In addition to eight mammalian orders, we detected an even higher diversity of birds, besides amphibians, reptiles, and even fish, from an average of 63 flies per locality. This remarkable diversity would be challenging to achieve using a single traditional method, such as camera traps or direct observation transects. Comparative studies have shown that iDNA metabarcoding detects greater species diversity than camera traps (e.g., Massey et al., 2021), emphasizing its ability to provide a comprehensive snapshot of biodiversity, particularly in environments where traditional methods are limited by factors such as animal behavior or habitat accessibility (Lee et al., 2023). Moreover, accumulation curves from previous studies (Calvignac-Spencer et al., 2013; Saranholi et al., 2024) suggest that larger sampling efforts could recover more taxa, as the detection of new mammal species increases rapidly until approximately 150 invertebrate specimens used as iDNA sources.

iDNA metabarcoding has been successfully employed to identify a wide range of mammal species using various invertebrate samplers, including carrion flies (Calvignac-Spencer et al., 2013; Rodgers et al., 2017), dung beetles (Drinkwater et al., 2021; Saranholi et al., 2024), leeches (Schnell et al., 2012), mosquitoes, and sand flies (Danabalan et al., 2023). In this study, we reaffirmed the utility of carrion flies (Calliphoridae and Sarcophagidae) and introduced horseflies (Tabanidae) as novel iDNA sources. For the first time, horseflies were shown to be effective samplers, with their advanced flight capabilities and host-seeking adaptations enabling the detection of elusive mammals and a broad diversity of vertebrates in tropical environments.

The collection seasons appeared to influence the abundance of flies sampled and, consequently, the number of vertebrate OTUs detected in each locality. Dry and cooler periods resulted in fewer flies and vertebrate OTUs, consistent with studies showing that seasonality and meteorological variables, such as temperature and humidity, significantly affect fly activity and abundance (Koller et al., 2019; Hennekeler et al., 2011). For instance, horsefly populations in tropical regions of Brazil peak during the rainy season (Krüger & Krolow, 2015). To maximize sampling efficiency in Neotropical locations, we suggest that fly collection for iDNA sampling be conducted during warmer and more humid periods whenever possible.

While a substantial proportion of Calliphoridae flies presented vertebrate sequences (40%), consistent with previous studies (Calvignac-Spencer et al., 2013), the same was not observed for Tabanidae and Sarcophagidae flies (11% and 9%, respectively). This may be attributed to the high efficiency of Calliphoridae flies in locating and exploiting carrion, facilitated by their acute olfactory senses. As primary colonizers of carrion, these flies often encounter tissues—and consequently DNA—in relatively intact conditions (Sontigun et al., 2018; Zajac et al., 2018). In contrast, Sarcophagidae flies may exhibit different feeding strategies, including coprophagy (feeding on feces) (Borghesan et al., 2018), which is a poor source of DNA (Chiou & Bergey, 2018). Regarding Tabanidae, their dual feeding strategy, with females being primarily hematophagous but also feeding on nectar for energy (Kniepert, 1980), likely explains the lower proportion of vertebrate sequences. Additionally, hematophagous insects digest blood meals relatively quickly, a process that may accelerate during warmer periods (Campos et al., 2004).

Understanding the host preferences of vertebrate-feeding flies provides valuable insights into their ecological roles. The flies analyzed in this study exhibit opportunistic host selection, lacking strict specificity and targeting hosts based on their availability in their environment. Horseflies (Tabanidae) play a dual role as vectors for pathogens and pollinators, feeding opportunistically on a variety of hosts, including mammals, birds, and specific flowers (Kniepert 1980; Vaduva, 2015). Our results expanded this understanding, revealing that they can also feed on amphibians, reptiles, and even fish. Carrion flies, known for their role in decomposing organic material and recycling nutrients, also demonstrated opportunistic feeding behaviors. Sarcophagidae flies detected a diversity of mammals and some amphibians, while Calliphoridae detected mammals, amphibians, and birds, supporting prior findings on their feeding strategies (Bermúdez et al., 2007; Junes et al., 2019).

Pooling flies before amplification has been shown to reduce the number of taxa detected in some studies (Rodgers et al., 2017). However, our results demonstrated that pooled samples yielded a larger number and different composition of vertebrate OTUs compared to individual samples. Given that Tabanidae and Sarcophagidae have a very low proportion of flies containing vertebrate sequences, pooling individuals from these families can be an effective cost-reduction strategy. Nevertheless, individual samples also contained exclusive OTUs, suggesting that a combination of pooled and individual samples, particularly for Calliphoridae flies, may further maximize vertebrate detection.

Knowing that cross-contamination and tag jumping are a significant concern in DNA metabarcoding studies, which can lead to misidentifications, we implemented several mitigation measures (Murray et al., 2015), including dedicated PCR preparation spaces, decontamination of surfaces and equipment, twin-tags for amplicon identification, post-PCR purification steps, and standardization of sample concentrations before pooling. Despite these precautions, approximately 13% of corrected reads were identified as cross-contamination from positive controls, likely caused by aerosolized amplicons and sample carry-over during multi-round PCR. This proportion was considerably higher than the 2.1–2.6% reported in previous studies (e.g., Schnell et al., 2015). Using a single-step PCR protocol for library preparation and fewer positive controls, preferably from non-native species, could reduce this issue (Murray et al., 2015).

The two-step PCR protocol for Illumina sequencing used in this study was chosen for its primer flexibility, potential to reduce template-specific biases, and cost-effectiveness (O'Donnell et al., 2016; Chen et al., 2021). However, besides increasing the risk of contamination (Seitz et al., 2015), this protocol can significantly reduce amplification efficiency and yield due to mismatches caused by tail sequences added to primers (Marquina et al., 2018). Tests with the same COI primers demonstrated a 65% reduction in efficiency when using the two-step PCR protocol compared to Sanger sequencing (Chaves et al., submitted). This suggests that the quantity and diversity of vertebrates detected in the iDNA of sampled flies could have been greater with a protocol that did not compromise amplification efficiency.

We carried out the first application of iDNA metabarcoding in the highly diverse Atlantic Forest and Cerrado biomes, along with additional locations in the Amazon. While studies in temperate ecosystems benefit from well-established reference databases and cost-effectiveness improvements (Abrams et al., 2019; Jackman et al., 2021; Danabalan, 2023), metabarcoding in the Neotropics is still hindered by limited research funding and incomplete reference databases, besides sheltering immense genetic and species diversity (Sales et al., 2020; Carvalho et al., 2022; Teixeira, 2023). Despite employing a customized reference database and primer pairs specifically adapted for Brazilian mammals, we encountered significant challenges in species-level identification in "real samples," with only one-fifth of mammalian OTUs successfully assigned to species. Addressing these limitations will require expanding reference databases through active sampling and sequencing efforts, particularly for COI, RNA12S, and RNA16S sequences of Brazilian mammals.

While the COI gene is widely recognized for its high taxonomic resolution and extensive reference database for vertebrate species, it is less commonly used in metabarcoding studies due to primer bias caused by its high sequence variability across species, besides the availability of universal primers for other genetic markers, such as RNA12S and RNA16S, which amplify a broader range of taxa (Riaz et al., 2011). By combining the high taxonomic resolution of COI with the broader taxonomic coverage of RNA12S and RNA16S—using specifically selected and adapted primer pairs for Brazilian mammals (Chaves et al., submitted)—we detected a total of 106 OTUs spanning a large diversity of vertebrates. While COI and RNA12S primarily detected mammals and birds, RNA16S contributed the most to taxonomic diversity, identifying a wide range of mammals, birds, amphibians, reptiles, and fish. Beyond its broader taxonomic coverage, the concomitant use of ribosomal genes likely enhanced vertebrate detection, as COI sequences accounted for only one-quarter of the detected OTUs, possibly due to their sensitivity to degradation, particularly in samples where DNA quality is often compromised (Ficetola et al., 2010).

Incorporating information about species distributions, particularly for birds and mammals, allowed for more reliable taxonomic assignments, although it limited identifications at the species level. Without this approach, all

amphibians, reptiles, and fishes had to be assigned to the family level or higher. Applying a similar strategy to these classes could enhance the reliability of their identifications. Additionally, comparing detected OTUs with existing species lists from each sampled locality, as done in previous studies (Rodgers et al., 2017; Saranholi et al., 2024), could improve the accuracy of taxonomic assignments and potentially increase the number of samples identified at the species level.

When combined with other sampling sources, such as water and other invertebrates, metabarcoding of horseflies and carrion flies can significantly enhance biodiversity monitoring and conservation efforts in Neotropical biomes. In long-term studies, this approach could enable the detection of elusive and rare taxa often missed by traditional methods, supporting the prioritization of conservation actions for vulnerable species and habitats (Fernandes et al., 2023). As highlighted by Ji et al. (2013), metabarcoding allows a shift from biodiversity indicators to direct measurements of total biodiversity, offering a more comprehensive view of ecosystem health. Integrating metabarcoding with complementary methods, such as camera traps and acoustic monitoring, can further strengthen biodiversity assessments by combining data on species distributions and interactions within ecosystems (Hoffmann et al., 2018).

CONCLUSION

This study demonstrated the potential of iDNA metabarcoding as a robust tool for surveying mammalian and broader vertebrate diversity in Neotropical biomes, including the highly diverse Atlantic Forest and Cerrado. By combining innovative sampling approaches using horseflies and carrion flies with tailored genetic markers and a customized reference database, we detected a wide range of taxa, encompassing elusive, endangered, and even exotic species. Despite challenges in species-level identification due to incomplete reference databases and methodological constraints, our findings highlight the capacity of iDNA metabarcoding to complement traditional biodiversity monitoring techniques, offering comprehensive and scalable insights into vertebrate communities. Addressing existing limitations, particularly by expanding reference databases and refining amplification protocols, will further enhance its utility in informing conservation strategies and supporting ecosystem health assessments across the Neotropics.

REFERENCES

Abrams, J. F., Hörig, L. A., Brozovic, R., Axtner, J., Crampton-Platt, A., Mohamed, A., Wong, S. T., Sollmann, R., Yu, D. W., & Wilting, A. (2019). Shifting up a gear with iDNA: From mammal detection events to standardized surveys. *Journal of Applied Ecology*, *56*(7), 1637–1648. https://doi.org/10.1111/1365-2664.13411

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arponen, A. (2012). Prioritizing species for conservation planning. *Biodiversity and Conservation*, *21*(4), 875–893. https://doi.org/10.1007/S10531-012-0242-1/FIGURES/2

Bermúdez, S. E., Espinosa, J. D., Cielo, A. B., Clavel, F., Subía, J., Barrios, S., & Medianero, E. (2007). Incidence of myiasis in Panama during the eradication of Cochliomyia hominivorax (Coquerel 1858, Diptera: Calliphoridae) (2002-2005). *Memórias Do Instituto Oswaldo Cruz*, *102*(6), 675–679. https://doi.org/10.1590/S0074-02762007005000074

Blahó, M., Egri, Á., Száz, D., Kriska, G., Åkesson, S., & Horváth, G. (2013). Stripes disrupt odor attractiveness to biting horseflies: Battle between ammonia, CO2, and color pattern for dominance in the sensory systems of host-seeking tabanids. *Physiology & Behavior*, *119*, 168–174. https://doi.org/10.1016/J.PHYSBEH.2013.06.013

Bogoni, J. A., Pires, J. S. R., Graipel, M. E., Peroni, N., & Peres, C. A. (2018). Wish you were here: How defaunated is the Atlantic Forest biome of its medium- to large-bodied mammal fauna? *PLOS ONE*, *13*(9), e0204515. https://doi.org/10.1371/JOURNAL.PONE.0204515

Borghesan, T. C., Campaner, M., Matsumoto, T. E., Espinosa, O. A., Razafindranaivo, V., Paiva, F., Carranza, J. C., Añez, N., Neves, L., Teixeira, M. M. G., & Camargo, E. P. (2018). Genetic diversity and phylogenetic relationships of coevolving symbiont-harboring insect trypanosomatids, and their neotropical dispersal by invader African blowflies (Calliphoridae). *Frontiers in Microbiology*, *9*(FEB), 293109. https://doi.org/10.3389/FMICB.2018.00131/BIBTEX

Brown, B. v. (2020). Sampling methods for adult flies (Diptera). In *Measuring Arthropod Biodiversity: A Handbook of Sampling Methods* (pp. 187–204). Springer International Publishing. https://doi.org/10.1007/978-3-030-53226-0_7

Calvignac-Spencer, S., Leendertz, F. H., Gilbert, M. T. P., & Schubert, G. (2013). An invertebrate stomach's view on vertebrate ecology: Certain invertebrates could be used as "vertebrate samplers" and deliver DNA-based information on many aspects of vertebrate ecology. *BioEssays*, *35*(11), 1004–1013. https://doi.org/10.1002/bies.201300060

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P. M., Metzger, S., Schubert, G., & Leendertz, F. H. (2013). Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology*, *22*(4), 915–924. https://doi.org/10.1111/mec.12183

Campos, I. T. N., Tanaka-Azevedo, A. M., & Tanaka, A. S. (2004a). Identification and characterization of a novel factor XIIa inhibitor in the hematophagous insect, Triatoma infestans (Hemiptera: Reduviidae). *FEBS Letters*, *577*(3), 512–516. https://doi.org/10.1016/J.FEBSLET.2004.10.052

Campos, I. T. N., Tanaka-Azevedo, A. M., & Tanaka, A. S. (2004b). Identification and characterization of a novel factor XIIa inhibitor in the hematophagous insect, Triatoma infestans (Hemiptera: Reduviidae). *FEBS Letters*, *577*(3), 512–516. https://doi.org/10.1016/J.FEBSLET.2004.10.052

Campos, I. T. N., Tanaka-Azevedo, A. M., & Tanaka, A. S. (2004c). Identification and characterization of a novel factor XIIa inhibitor in the hematophagous insect, Triatoma infestans (Hemiptera: Reduviidae). *FEBS Letters*, *577*(3), 512–516. https://doi.org/10.1016/j.febslet.2004.10.052

Carvalho, C. S., de Oliveira, M. E., Rodriguez-Castro, K. G., Saranholi, B. H., & Galetti, P. M. (2022). Efficiency of eDNA and iDNA in assessing vertebrate diversity and its abundance. *Molecular Ecology Resources*, *22*(4), 1262–1273. https://doi.org/10.1111/1755-0998.13543

Chen, K. H., Longley, R., Bonito, G., & Liao, H. L. (2021). A two-step PCR protocol enabling flexible primer choice and high sequencing yield for illumina miseq meta-barcoding. In *Agronomy* (Vol. 11, Issue 7). MDPI AG. https://doi.org/10.3390/agronomy11071274

Chiou, K. L., & Bergey, C. M. (2018). Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. *Scientific Reports 2018 8:1*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-20427-9

Clarke, L. J., Beard, J. M., Swadling, K. M., & Deagle, B. E. (2017). Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and Evolution*, *7*(3), 873–883. https://doi.org/10.1002/ECE3.2667

Danabalan, R., Planillo, A., Butschkau, S., Deeg, S., Pierre, G., Thion, C., Calvignac-Spencer, S., Kramer-Schadt, S., & Mazzoni, C. (2023). Comparison of mosquito and fly-derived DNA as a tool for sampling vertebrate biodiversity in suburban forests in Berlin, Germany. *Environmental DNA*, *5*(3), 476–487. https://doi.org/10.1002/edn3.398

de Piacentini, V. Q., Aleixo, A., Agne, C. E., Maurício, G. N., Pacheco, J. F., Bravo, G. A., Brito, G. R. R., Naka, L. N., Olmos, F., Posso, S., Silveira, L. F., Betini, G. S., Carrano, E., Franz, I., Lees, A. C., Lima, L. M., Pioli, D., Schunck, F., do Amaral, F. R., … Cesari, E. (2015). Annotated checklist of the birds of Brazil by the Brazilian Ornithological Records Committee / Lista comentada das aves do Brasil pelo Comitê Brasileiro de Registros Ornitológicos. *Revista Brasileira de Ornitologia 2015 23:2*, *23*(2), 91–298. https://doi.org/10.1007/BF03544294

Donlan, C. J. (Ed.). (2015). *Proactive strategies for protecting species: pre-listing conservation and the Endangered Species Act*. Univ of California Press.

Drinkwater, R., Williamson, J., Clare, E. L., Chung, A. Y. C., Rossiter, S. J., & Slade, E. (2021). Dung beetles as samplers of mammals in Malaysian Borneo-a test of high throughput metabarcoding of iDNA. *PeerJ*, *9*, e11897. https://doi.org/10.7717/PEERJ.11897/SUPP-6

Fernandes, K., Bateman, P. W., Saunders, B. J., Bunce, M., Bohmann, K., & Nevill, P. (2023). Use of carrion fly iDNA metabarcoding to monitor invasive and native mammals. *Conservation Biology*, *37*(5), e14098. https://doi.org/10.1111/COBI.14098

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., & Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. *BMC Genomics*, *11*(1). https://doi.org/10.1186/1471-2164-11-434

Galetti, M., Brocardo, C. R., Begotti, R. A., Hortenci, L., Rocha-Mendes, F., Bernardo, C. S. S., Bueno, R. S., Nobre, R., Bovendorp, R. S., Marques, R. M., Meirelles, F., Gobbo, S. K., Beca, G., Schmaedecke, G., & Siqueira, T. (2017). Defaunation and biomass collapse of mammals in the largest Atlantic forest remnant. *Animal Conservation*, *20*(3), 270–281. https://doi.org/10.1111/ACV.12311

Gogarten, J. F., Hoffmann, C., Arandjelovic, M., Sachse, A., Merkel, K., Dieguez, P., Agbor, A., Angedakin, S., Brazzola, G., Jones, S., Langergraber, K. E., Lee, K., Marrocoli, S., Murai, M., Sommer, V., Kühl, H., Leendertz, F. H., & Calvignac-Spencer, S. (2020). Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity. *Environmental DNA*, *2*(1), 63–76. https://doi.org/10.1002/edn3.46

Haile, J., Froese, D. G., MacPhee, R. D. E., Roberts, R. G., Arnold, L. J., Reyes, A. v., Rasmussen, M., Nielsen, R., Brook, B. W., Robinson, S., Demuro, M., Gilbert, M. T. P., Munch, K., Austin, J. J., Cooper, A., Barnes, I., Möller, P., & Willerslev, E. (2009). Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(52), 22352–22357. https://doi.org/10.1073/pnas.0912510106

Hajibabaei, M. (2012). The golden age of DNA metasystematics. In *Trends in Genetics* (Vol. 28, Issue 11, pp. 535–537). https://doi.org/10.1016/j.tig.2012.08.001

Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J., & Knight, R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods 2008 5:3*, *5*(3), 235–237. https://doi.org/10.1038/nmeth.1184

Hoffmann, C., Merkel, K., Sachse, A., Rodríguez, P., Leendertz, F. H., & Calvignac-Spencer, S. (2018a). Blow flies as urban wildlife sensors. *Molecular Ecology Resources*, *18*(3), 502–510. https://doi.org/10.1111/1755-0998.12754

Hoffmann, C., Merkel, K., Sachse, A., Rodríguez, P., Leendertz, F. H., & Calvignac-Spencer, S. (2018b). Blow flies as urban wildlife sensors. *Molecular Ecology Resources*, *18*(3), 502–510. https://doi.org/10.1111/1755-0998.12754

Hoffmann, C., Merkel, K., Sachse, A., Rodríguez, P., Leendertz, F. H., & Calvignac-Spencer, S. (2018c). Blow flies as urban wildlife sensors. *Molecular Ecology Resources*, *18*(3), 502–510. https://doi.org/10.1111/1755-0998.12754

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, *44*(9), 1259–1269. https://doi.org/10.1111/ECOG.05926

IUCN 2024. The IUCN Red List of Threatened Species. Version 2024-2. <https://www.iucnredlist.org>

Jackman, J. M., Benvenuto, C., Coscia, I., Oliveira Carvalho, C., Ready, J. S., Boubli, J. P., Magnusson, W. E., McDevitt, A. D., & Guimarães Sales, N. (2021). eDNA in a bottleneck: Obstacles to fish metabarcoding studies in megadiverse freshwater systems. *Environmental DNA*, *3*(4), 837–849. https://doi.org/10.1002/EDN3.191

Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), E2603–E2610. https://doi.org/10.1073/pnas.1302251110

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C., Wang, X., Levi, T., Lott, M., … Yu, D. W. (2013). Reliable, verifiable, and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), 1245–1257. https://doi.org/10.1111/ele.12162

Junes, K., Ruiz, J., & Quispitupac, E. (2019). Flesh-fly myiasis (Diptera: Sarcophagidae) in Dendropsophus schubarti (Anura: Hylidae) from Peru. *Phyllomedusa*, *18*(2), 277–281. https://doi.org/10.11606/issn.2316-9079.v18i2p277-281

Keck, F., Couton, M., & Altermatt, F. (2023). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Molecular Ecology Resources*, *23*(4), 742–755. https://doi.org/10.1111/1755-0998.13746

Kniepert, F. W. (1980). Blood-feeding and nectar-feeding in adult tabanidae (Diptera). *Oecologia*, *46*(1), 125–129. https://doi.org/10.1007/BF00346976/METRICS

Kocher, A., de Thoisy, B., Catzeflis, F., Valière, S., Bañuls, A. L., & Murienne, J. (2017). iDNA screening: Disease vectors as vertebrate samplers. *Molecular Ecology*, *26*(22), 6478–6486. https://doi.org/10.1111/mec.14362

Krüger, R. F., & Krolow, T. K. (2015). Seasonal patterns of horse fly richness and abundance in the Pampa biome of southern Brazil. *Journal of Vector Ecology*, *40*(2), 364–372. https://doi.org/10.1111/JVEC.12175

Lee, P., He, T., Dong, M., Huang, Q., Zhou, X., Liao, J., Chen, X., Wu, X., Wee, S. L., & Chen, J. (2023). Comparative Study on Blowfly-Derived DNA and Camera Trapping in Assessing Mammalian Diversity in Subtropical Forests. *Forests*, *14*(11). https://doi.org/10.3390/f14112180

Lee, P. S., Sing, K. W., & Wilson, J. J. (2015). Reading mammal diversity from flies: The persistence period of amplifiable mammal mtDNA in blowfly guts (Chrysomya megacephala) and a new DNA mini-barcode target. *PLoS ONE*, *10*(4). https://doi.org/10.1371/journal.pone.0123871

Lynggaard, C., Nielsen, M., Santos-Bay, L., Gastauer, M., Oliveira, G., & Bohmann, K. (2019). Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests. *Environmental DNA*, *1*(4), 329–341. https://doi.org/10.1002/edn3.34

Marquina, D., Esparza-Salas, R., Roslin, T., & Ronquist, F. (2019). Establishing arthropod community composition using metabarcoding: Surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. *Molecular Ecology Resources*, *19*(6), 1516–1530. https://doi.org/10.1111/1755-0998.13071

Massey, A. L., Bronzoni, R. V. de M., da Silva, D. J. F., Allen, J. M., de Lázari, P. R., dos Santos-Filho, M., Canale, G. R., Bernardo, C. S. S., Peres, C. A., & Levi, T. (2022). Invertebrates for vertebrate

biodiversity monitoring: Comparisons using three insect taxa as iDNA samplers. *Molecular Ecology Resources*, *22*(3), 962–977. https://doi.org/10.1111/1755-0998.13525

Meyerson, L. A., Pauchard, A., Brundu, G., Carlton, J. T., Hierro, J. L., Kueffer, C., Pandit, M. K., Pyšek, P., Richardson, D. M., & Packer, J. G. (2022). Moving Toward Global Strategies for Managing Invasive Alien Species. *Global Plant Invasions*, 331–360. https://doi.org/10.1007/978-3-030-89684-3_16

Murray, D. C., Coghlan, M. L., & Bunce, M. (2015). From benchtop to desktop: Important considerations when designing amplicon sequencing workflows. *PLoS ONE*, *10*(4). https://doi.org/10.1371/journal.pone.0124671

Myers, N., Mittermeier[2], R. A., Mittermeier[2], C. G., da Fonseca[3], G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. In *NATURE* | (Vol. 403). www.nature.com

O'donnell, J. L., Kelly, R. P., Lowell, N. C., & Port, J. A. (2016). Indexed PCR Primers Induce Template-Specific Bias in Large-Scale DNA Sequencing Studies. *PLOS ONE*, *11*(3), e0148698. https://doi.org/10.1371/JOURNAL.PONE.0148698

Oliveira, R. R. M., Silva, R., Nunes, G. L., & Oliveira, G. (2021). *PIMBA: a PIpeline for MetaBarcoding Analysis*. https://doi.org/10.1101/2021.03.23.436646

Paglia, A. P., da Fonseca, G. A. B., Rylands, A. B., Herrmann, G., Aguiar, L. M. S., Chiarello, A. G., Leite, Y. L. R., Costa, L. P., Siciliano, S., Cecília, M., Kierulff, M., Mendes, S. L., Da, V., Tavares, C., Mittermeier, R. A., & Patton, J. L. (2012). *OCCASIONAL PAPERS IN CONSERVATION BIOLOGY Lista Anotada dos Mamíferos do Brasil 2ª Edição Annotated Checklist of Brazilian Mammals 2 nd Edition*.

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). EcoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, *39*(21). https://doi.org/10.1093/nar/gkr732

Rodgers, T. W., Xu, C. C. Y., Giacalone, J., Kapheim, K. M., Saltonstall, K., Vargas, M., Yu, D. W., Somervuo, P., McMillan, W. O., & Jansen, P. A. (2017). Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community. *Molecular Ecology Resources*, *17*(6), e133–e145. https://doi.org/10.1111/1755-0998.12701

Sales, N. G., McKenzie, M. B., Drake, J., Harper, L. R., Browett, S. S., Coscia, I., Wangensteen, O. S., Baillie, C., Bryce, E., Dawson, D. A., Ochu, E., Hänfling, B., Lawson Handley, L., Mariani, S., Lambin, X., Sutherland, C., & McDevitt, A. D. (2020). Fishing for mammals: Landscape-level monitoring of terrestrial and semi-aquatic communities using eDNA from riverine systems. *Journal of Applied Ecology*, *57*(4), 707–716. https://doi.org/10.1111/1365-2664.13592

Saranholi, B. H., Rodriguez-Castro, K. G., Carvalho, C. S., Chahad-Ehlers, S., Gestich, C. C., Andrade, S. C. S., Freitas, P. D., & Galetti, P. M. (2023). Comparing iDNA from mosquitoes and flies to survey mammals in a semi-controlled Neotropical area. *Molecular Ecology Resources*, *23*(8), 1790–1799. https://doi.org/10.1111/1755-0998.13851

Saranholi, B. H., França, F. M., Vogler, A. P., Barlow, J., Vaz de Mello, F. Z., Maldaner, M. E., Carvalho, E., Gestich, C. C., Howes, B., Banks-Leite, C., & Galetti, P. M. (2024). Testing and optimizing metabarcoding of iDNA from dung beetles to sample mammals in the hyperdiverse Neotropics. *Molecular Ecology Resources*, *24*(5). https://doi.org/10.1111/1755-0998.13961

Schipper, J., Chanson, J. S., Chiozza, F., Cox, N. A., Hoffmann, M., Katariya, V., Lamoreux, J., Rodrigues, A. S. L., Stuart, S. N., Temple, H. J., Baillie, J., Boitani, L., Lacher, T. E., Mittermeier, R. A., Smith, A. T., Absolon, D., Aguiar, J. M., Amori, G., Bakkour, N., … Young, B. E. (2008). The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science*, *322*(5899), 225–230. https://doi.org/10.1126/SCIENCE.1165115/SUPPL_FILE/SCHIPPER.SOM.PDF

Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R. D., Willerslev, E., Bertelsen, M. F., & Gilbert, M. T. P. (2012). Screening mammal biodiversity using DNA from leeches. In *Current Biology* (Vol. 22, Issue 8). https://doi.org/10.1016/j.cub.2012.02.058

Seitz, V., Schaper, S., Dröge, A., Lenze, D., Hummel, M., & Hennig, S. (2015). A new method to prevent carry-over contaminations in two-step PCR NGS library preparations. *Nucleic Acids Research*, *43*(20), e135–e135. https://doi.org/10.1093/NAR/GKV694

Skvarla, M. J., Larson, J. L., Fisher, J. R., & Dowling, A. P. G. (2021). A Review of Terrestrial and Canopy Malaise Traps. In *Annals of the Entomological Society of America* (Vol. 114, Issue 1, pp. 27–47). Entomological Society of America. https://doi.org/10.1093/aesa/saaa044

Somervuo, P., Koskela, S., Pennanen, J., Henrik Nilsson, R., & Ovaskainen, O. (2016). Unbiased probabilistic taxonomic classification for DNA barcoding. *Bioinformatics*, *32*(19), 2920–2927. https://doi.org/10.1093/bioinformatics/btw346

Sontigun, N., Sukontason, K. L., Klong-Klaew, T., Sanit, S., Samerjai, C., Somboon, P., Thanapornpoonpong, S. N., Amendt, J., & Sukontason, K. (2018). Bionomics of the oriental latrine fly Chrysomya megacephala (Fabricius) (Diptera: Calliphoridae): temporal fluctuation and reproductive potential. *Parasites and Vectors*, *11*(1), 1–12. https://doi.org/10.1186/S13071-018-2986-2/FIGURES/9

Teixeira, D. F., Hilário, H. O., Santos, G. B., & Carvalho, D. C. (2023). DNA metabarcoding assessment of Neotropical ichthyoplankton communities is marker-dependent. *Ecology and Evolution*, *13*(10), e10649. https://doi.org/10.1002/ECE3.10649

Thomas, A. W., & Gooding, R. H. (1976). Digestive Processes Of Hematophagous Insects: VIII. Estimation of meal size and demonstration of trypsin in horse flies and deer flies (Diptera: Tabanidae). *Journal of Medical Entomology*, *13*(2), 131–136. https://doi.org/10.1093/JMEDENT/13.2.131

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., Orlando, L., & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, *21*(11), 2565–2573. https://doi.org/10.1111/j.1365-294X.2011.05418.x

Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., Murata, K., Nikaido, M., Sado, T., Sato, Y., Takeshita, M., Iwasaki, W., Yamanaka, H., Kondoh, M., & Miya, M. (2017). Environmental DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources*, *17*(6), e63–e75. https://doi.org/10.1111/1755-0998.12690

Vaduva, G. (2015). A study on bloodsucking Tabanidae and Stomoxys calcitrans (Diptera) attacking horses and cows in Northern Scania, Sweden. *Journal of Biology and Life Science*, *7*(1), 19–32. https://doi.org/10.5296/JBLS.V7I1.8611

van Hennekeler, K., Jones, R. E., Skerratt, L. F., Muzari, M. O., & Fitzpatrick, L. A. (2011). Meteorological effects on the daily activity patterns of tabanid biting flies in northern Queensland, Australia. *Medical and Veterinary Entomology*, *25*(1), 17–24. https://doi.org/10.1111/J.1365-2915.2010.00900.X

Wilson, J. J. (2012). DNA Barcodes for Insects. *Methods in Molecular Biology*, *858*, 17–46. https://doi.org/10.1007/978-1-61779-591-6_3

Zajac, B. K., Amendt, J., Verhoff, M. A., & Zehner, R. (2018). Dating Pupae of the Blow Fly Calliphora vicina Robineau–Desvoidy 1830 (Diptera: Calliphoridae) for Post Mortem Interval—Estimation: Validation of Molecular Age Markers. *Genes 2018, Vol. 9, Page 153*, *9*(3), 153. https://doi.org/10.3390/GENES9030153

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., Chariton, A. A., Creer, S., Coissac, E., Deagle, B. E., de Barba, M., Dickie, I. A., Dumbrell, A. J., Ficetola, G. F., Fierer, N., Fumagalli, L., Gilbert, M. T. P., Jarman, S., Jumpponen, A., … Taberlet, P. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, *28*(8), 1857–1862. https://doi.org/10.1111/MEC.15060

4. Considerações Finais

Este trabalho contribuiu para o avanço do conhecimento e das metodologias relacionadas ao *metabarcoding* de iDNA aplicado à identificação de mamíferos neotropicais. No primeiro capítulo, a construção de um banco de referência customizado, cobrindo 72% das espécies de mamíferos brasileiros, e a avaliação de marcadores genéticos e primers frequentemente utilizados na literatura permitiram otimizar a precisão das identificações taxonômicas. A combinação dos marcadores COI, RNA12S e RNA16S mostrou-se uma abordagem eficaz, com potencial para superar desafios impostos pela diversidade de espécies nos biomas brasileiros.

No segundo capítulo, a aplicação prática dessa metodologia em amostras reais ampliou o entendimento sobre a biodiversidade dos biomas Amazônia, Mata Atlântica e Cerrado. A inclusão de mutucas (Tabanidae) como amostradores de iDNA, ao lado de moscas varejeiras (Calliphoridae e Sarcophagidae), possibilitou a detecção de 106 unidades taxonômicas operacionais (OTUs), abrangendo não apenas várias ordens de mamíferos, mas também diferentes grupos de vertebrados, incluindo uma espécie ameaçada e uma exótica. Apesar de desafios como lacunas nos bancos de referência, sazonalidade e contaminação cruzada, os resultados destacam o potencial do *metabarcoding* como ferramenta complementar aos métodos tradicionais de levantamento da biodiversidade.

As contribuições deste estudo são relevantes, oferecendo diretrizes práticas para a aplicação do *metabarcoding* na região neotropical. Os resultados demonstram a importância de investir em bancos de referência mais abrangentes e em metodologias adaptadas à diversidade neotropical, visando ampliar a eficácia e aplicabilidade do *metabarcoding* em programas de biomonitoramento e conservação. Estudos futuros devem se concentrar no aumento contínuo da representatividade de espécies neotropicais nos bancos de referência, bem como no aprimoramento de protocolos mais eficientes para a detecção de espécies.

Esta tese reforça a necessidade de integrar abordagens inovadoras, como o *metabarcoding*, a estratégias de conservação e manejo sustentável, especialmente em *hotspots* de biodiversidade. Ao permitir a detecção de uma ampla gama de espécies com o mínimo de distúrbio e esforço, o *metabarcoding* pode ser integrado a programas rotineiros de biomonitoramento, otimizando a coleta e a análise de dados e permitindo uma análise cada vez mais abrangente da biodiversidade total e da saúde dos ecossistemas. Assim, a expansão dessas metodologias poderá oferecer suporte a políticas públicas voltadas à preservação dos biomas brasileiros.

As listas de espécies obtidas por meio de *metabarcoding* também podem contribuir para estudos biogeográficos, possibilitando a avaliação da distribuição de espécies em diferentes habitats e regiões geográficas. A análise de eDNA de diversas fontes pode permitir a identificação de espécies em áreas de difícil acesso ou que apresentam desafios para levantamento por métodos convencionais. Com o contínuo aprimoramento das metodologias e a ampliação dos bancos de referência, espera-se que o papel do *metabarcoding* em estudos de biodiversidade cresça, proporcionando insights valiosos sobre as complexidades da vida na Terra.

REFERÊNCIAS GERAIS

Allen, M. C., Kwait, R., Vastano, A., Kisurin, A., Zoccolo, I., Jaffe, B. D., Angle, J. C., Maslo, B., & Lockwood, J. L. (2022). *Sampling environmental DNA from trees and soil to detect cryptic arboreal mammals*. https://doi.org/10.21203/RS.3.RS-1862862/V1

Andújar, C., Arribas, P., Gray, C., Bruce, C., Woodward, G., Yu, D. W., & Vogler, A. P. (2018). Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. *Molecular Ecology*, *27*(1), 146–166. https://doi.org/10.1111/mec.14410

Axtner, J., Crampton-Platt, A., Hörig, L. A., Mohamed, A., Xu, C. C., Yu, D. W., & Wilting, A. (2019). An efficient and robust laboratory workflow and tetrapod database for larger scale environmental DNA studies. *GigaScience*, 8(4), giz029. https://doi.org/10.1093/gigascience/giz029

Belle, C. C., Stoeckle, B. C., & Geist, J. (2019). Taxonomic and geographical representation of freshwater environmental DNA research in aquatic conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *29*(11), 1996–2009. https://doi.org/10.1002/AQC.3208

Bernard, H., Ahmad, A. H., Brodie, J., Giordano, A. J., Lakim, M., Amat, R., ... & Sinun, W. (2013). Camera-trapping survey of mammals in and around Imbak Canyon Conservation Area in Sabah, Malaysian Borneo. *Raffles Bulletin of Zoology*, *61*(2).

Blahó, M., Egri, Á., Száz, D., Kriska, G., Åkesson, S., & Horváth, G. (2013). Stripes disrupt odour attractiveness to biting horseflies: Battle between ammonia, CO2, and colour pattern for dominance in the sensory systems of host-seeking tabanids. *Physiology & Behavior*, *119*, 168–174. https://doi.org/10.1016/J.PHYSBEH.2013.06.013

Brown, B. V. (2020). Sampling methods for adult flies (Diptera). In *Measuring Arthropod Biodiversity: A Handbook of Sampling Methods* (pp. 187–204). Springer International Publishing. https://doi.org/10.1007/978-3-030-53226-0_7

Calvignac-Spencer, S., Leendertz, F. H., Gilbert, M. T. P., & Schubert, G. (2013). An invertebrate stomach's view on vertebrate ecology: Certain invertebrates could be used as "vertebrate samplers" and deliver DNA-based information on many aspects of vertebrate ecology. *BioEssays*, *35*(11), 1004–1013. https://doi.org/10.1002/bies.201300060

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P. M., Metzger, S., Schubert, G., & Leendertz, F. H. (2013). Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology*, *22*(4), 915–924. https://doi.org/10.1111/mec.12183

Carvalho, C. S., de Oliveira, M. E., Rodriguez-Castro, K. G., Saranholi, B. H., & Galetti, P. M. (2022). Efficiency of eDNA and iDNA in assessing vertebrate diversity and its abundance. *Molecular Ecology Resources*, *22*(4), 1262–1273. https://doi.org/10.1111/1755-0998.13543

Crisci, J. v. (2006). One-Dimensional Systematist: Perils in a Time of Steady Progress. *Systematic Botany*, *31*(1), 217–221. http://www.jstor.org/stable/25064142

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. In *Molecular Ecology* (Vol. 26, Issue 21, pp. 5872–5895). Blackwell Publishing Ltd. https://doi.org/10.1111/mec.14350

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Lucas, A. G., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., … Zayas, C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, *366*(6471). https://doi.org/10.1126/SCIENCE.AAX3100/SUPPL_FILE/AAX3100-DIAZ-SM.PDF

Dirzo, R., & Raven, P. H. (2003). Global state of biodiversity and loss. *Annual review of Environment and Resources*, *28*(1), 137-167. https://doi.org/10.1146/annurev.energy.28.050302.105532

Felsenstein, J. (2004). Inferring phylogenies. 664. https://global.oup.com/academic/product/inferring-phylogenies-9780878931774

Ficetola, G. F., Boyer, F., Valentini, A., Bonin, A., Meyer, A., Dejean, T., Gaboriaud, C., Usseglio-Polatera, P., & Taberlet, P. (2021). Comparison of markers for the monitoring of freshwater benthic biodiversity through DNA metabarcoding. *Molecular Ecology*, *30*(13), 3189–3202. https://doi.org/10.1111/MEC.15632

Freeland, J. R. (2017). The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA. Genome, 60(4), 358-374. https://doi.org/10.1139/gen-2016-0100

Glover-Kapfer, P., Soto-Navarro, C. A., & Wearn, O. R. (2019). Camera-trapping version 3.0: current constraints and future priorities for development. *Remote Sensing in Ecology and Conservation*, *5*(3), 209–223. https://doi.org/10.1002/rse2.106

Gogarten, J. F., Hoffmann, C., Arandjelovic, M., Sachse, A., Merkel, K., Dieguez, P., Agbor, A., Angedakin, S., Brazzola, G., Jones, S., Langergraber, K. E., Lee, K., Marrocoli, S., Murai, M., Sommer, V., Kühl, H., Leendertz, F. H., & Calvignac-Spencer, S. (2020). Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity. *Environmental DNA*, *2*(1), 63–76. https://doi.org/10.1002/edn3.46

Green, S. J., Venkatramanan, R., & Naqib, A. (2015). Deconstructing the polymerase chain reaction: understanding and correcting bias associated with primer degeneracies and primer-template mismatches. PloS one, 10(5), e0128122. https://doi.org/10.1371/journal.pone.0128122

Haile, J., Froese, D. G., MacPhee, R. D. E., Roberts, R. G., Arnold, L. J., Reyes, A. v., Rasmussen, M., Nielsen, R., Brook, B. W., Robinson, S., Demuro, M., Gilbert, M. T. P., Munch, K., Austin, J. J., Cooper, A., Barnes, I., Möller, P., & Willerslev, E. (2009). Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(52), 22352–22357. https://doi.org/10.1073/pnas.0912510106

Hajibabaei, M., Porter, T. M., Wright, M., & Rudar, J. (2019). COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLoS One, 14(9), e0220953. https://doi.org/10.1371/journal.pone.0220953

Hebert, P. D., Ratnasingham, S., & De Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(suppl_1), S96-S99. https://doi.org/10.1098/rsbl.2003.0025

Hillis, D. M. (1987). MOLECULAR VERSUS MORPHOLOGICAL APPROACHES TO SYSTEMATICS. In *Source: Annual Review of Ecology and Systematics* (Vol. 18).

Ilunga, J., Paul, M., Daniel, G., Robitaille, J. F., & Mallory, F. F. (2020). Dna yields and pcr amplification success using degraded animal corpses. American Journal of Biochemistry and Biotechnology, 16(1), 60-69. https://doi.org/10.3844/ajbbsp.2020.60.69

IUCN 2024. The IUCN Red List of Threatened Species. Version 2024-2. <https://www.iucnredlist.org>

Jackman, J. M., Benvenuto, C., Coscia, I., Oliveira Carvalho, C., Ready, J. S., Boubli, J. P., Magnusson, W. E., McDevitt, A. D., & Guimarães Sales, N. (2021). eDNA in a bottleneck: Obstacles to fish metabarcoding studies in megadiverse freshwater systems. *Environmental DNA*, *3*(4), 837–849. https://doi.org/10.1002/EDN3.191

Keck, F., Couton, M., & Altermatt, F. (2023). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Molecular Ecology Resources*, *23*(4), 742–755. https://doi.org/10.1111/1755-0998.13746

Kent, R. J. (2009). Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. Molecular ecology resources, 9(1), 4-18. https://doi.org/10.1111/j.1755-0998.2008.02469.x

Kniepert, F. W. (1980). Blood-feeding and nectar-feeding in adult tabanidae (Diptera). *Oecologia*, *46*(1), 125–129. https://doi.org/10.1007/BF00346976/METRICS

Kocher, A., de Thoisy, B., Catzeflis, F., Valière, S., Bañuls, A. L., & Murienne, J. (2017). iDNA screening: Disease vectors as vertebrate samplers. *Molecular Ecology*, *26*(22), 6478–6486. https://doi.org/10.1111/mec.14362

Larsen, B. B., Miller, E. C., Rhodes, M. K., & Wiens, J. J. (2017). Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life. *The Quarterly Review of Biology*, *92*(3), 229-265.

Lee, P., He, T., Dong, M., Huang, Q., Zhou, X., Liao, J., Chen, X., Wu, X., Wee, S. L., & Chen, J. (2023). Comparative Study on Blowfly-Derived DNA and Camera Trapping in Assessing Mammalian Diversity in Subtropical Forests. *Forests*, *14*(11). https://doi.org/10.3390/f14112180

Lynggaard, C., Nielsen, M., Santos-Bay, L., Gastauer, M., Oliveira, G., & Bohmann, K. (2019). Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests. *Environmental DNA*, *1*(4), 329–341. https://doi.org/10.1002/edn3.34

Martínez-De la Puente, J., Méndez, M., Ruiz, S., Godoy, J. A., Soriguer, R. C., & Figuerola, J. (2015). Individual identification of endangered species using mosquito blood meals: a proof-of-concept study in Iberian lynx. *Parasitology Research*, *114*(4), 1607–1610. https://doi.org/10.1007/s00436-015-4343-0

Massey, A. L., Bronzoni, R. V. de M., da Silva, D. J. F., Allen, J. M., de Lázari, P. R., dos Santos-Filho, M., Canale, G. R., Bernardo, C. S. S., Peres, C. A., & Levi, T. (2022). Invertebrates for vertebrate biodiversity monitoring: Comparisons using three insect taxa as iDNA samplers. *Molecular Ecology Resources*, *22*(3), 962–977. https://doi.org/10.1111/1755-0998.13525

Miller, S. E., Hausmann, A., Hallwachs, W., & Janzen, D. H. (2016). Advancing taxonomy and bioinventories with DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1702). https://doi.org/10.1098/RSTB.2015.0339

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? *PLOS Biology*, *9*(8), e1001127. https://doi.org/10.1371/JOURNAL.PBIO.1001127

Mota, T. F. M., Fabrin, T. M. C., Gasques, L. S., Filho, H. O., Prioli, A. J., & Prioli, S. M. A. P. (2018). Extraction of dna from micro-tissue for bat species identification. Mitochondrial DNA Part B, 3(2), 758-762. https://doi.org/10.1080/23802359.2018.1484261

Pereira, J. C., Chaves, R., Bastos, E., Leitão, A., & Guedes-Pinto, H. (2011). An efficient method for genomic dna extraction from different molluscs species. International Journal of Molecular Sciences, 12(11), 8086-8095. https://doi.org/10.3390/ijms12118086

Primmer, C. R., Møller, A. P., & Ellegren, H. (1996). A wide-range survey of cross-species microsatellite amplification in birds. *Molecular ecology*, *5*(3), 365-378.

Rodgers, T. W., Xu, C. C. Y., Giacalone, J., Kapheim, K. M., Saltonstall, K., Vargas, M., Yu, D. W., Somervuo, P., McMillan, W. O., & Jansen, P. A. (2017). Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community. *Molecular Ecology Resources*, *17*(6), e133–e145. https://doi.org/10.1111/1755-0998.12701

Saranholi, B. H., Rodriguez-Castro, K. G., Carvalho, C. S., Chahad-Ehlers, S., Gestich, C. C., Andrade, S. C. S., Freitas, P. D., & Galetti, P. M. (2023). Comparing iDNA from mosquitoes and flies to survey mammals in a semi-controlled Neotropical area. *Molecular Ecology Resources*, *23*(8), 1790–1799. https://doi.org/10.1111/1755-0998.13851

Saranholi, B. H., França, F. M., Vogler, A. P., Barlow, J., Vaz de Mello, F. Z., Maldaner, M. E., Carvalho, E., Gestich, C. C., Howes, B., Banks-Leite, C., & Galetti, P. M. (2024). Testing and optimizing metabarcoding of iDNA from dung beetles to sample mammals in the hyperdiverse Neotropics. *Molecular Ecology Resources*, *24*(5). https://doi.org/10.1111/1755-0998.13961

Schipper, J., Chanson, J. S., Chiozza, F., Cox, N. A., Hoffmann, M., Katariya, V., Lamoreux, J., Rodrigues, A. S. L., Stuart, S. N., Temple, H. J., Baillie, J., Boitani, L., Lacher, T. E., Mittermeier, R. A., Smith, A. T., Absolon, D., Aguiar, J. M., Amori, G., Bakkour, N., … Young, B. E. (2008). The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science*, *322*(5899), 225–230. https://doi.org/10.1126/SCIENCE.1165115/SUPPL_FILE/SCHIPPER.SOM.PDF

Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R. D., Willerslev, E., Bertelsen, M. F., & Gilbert, M. T. P. (2012). Screening mammal biodiversity using DNA from leeches. In *Current Biology* (Vol. 22, Issue 8). https://doi.org/10.1016/j.cub.2012.02.058

Skvarla, M. J., Larson, J. L., Fisher, J. R., & Dowling, A. P. G. (2021). A Review of Terrestrial and Canopy Malaise Traps. In *Annals of the Entomological Society of America* (Vol. 114, Issue 1, pp. 27–47). Entomological Society of America. https://doi.org/10.1093/aesa/saaa044

Snaddon, J., Petrokofsky, G., Jepson, P., & Willis, K. J. (2013). Biodiversity technologies: tools as change agents. *Biology Letters*, *9*(1). https://doi.org/10.1098/RSBL.2012.1029

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Teixeira, D. F., Hilário, H. O., Santos, G. B., & Carvalho, D. C. (2023). DNA metabarcoding assessment of Neotropical ichthyoplankton communities is marker-dependent. *Ecology and Evolution*, *13*(10), e10649. https://doi.org/10.1002/ECE3.10649

Teletchea, F. (2016). Systematics and Aquaculture: What Could They Bring to Each Other? *Journal of Life Sciences*, *10*(5). https://doi.org/10.17265/1934-7391/2016.05.006

Tessler, M., Galen, S. C., DeSalle, R., & Schierwater, B. (2022). Let's end taxonomic blank slates with molecular morphology. *Frontiers in Ecology and Evolution*, *10*, 1016412. https://doi.org/10.3389/FEVO.2022.1016412/BIBTEX

Tzafesta, E., Saccomanno, B., Zangaro, F., Vadrucci, M. R., Specchia, V., & Pinna, M. (2022). DNA Barcode Gap Analysis for Multiple Marker Genes for Phytoplankton Species Biodiversity in Mediterranean Aquatic Ecosystems. *Biology*, *11*(9), 1277. https://doi.org/10.3390/BIOLOGY11091277/S1

Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., Murata, K., Nikaido, M., Sado, T., Sato, Y., Takeshita, M., Iwasaki, W., Yamanaka, H., Kondoh, M., & Miya, M. (2017). Environmental DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources*, *17*(6), e63–e75. https://doi.org/10.1111/1755-0998.12690

Vaduva, G. (2015). A Study on Bloodsucking Tabanidae and Stomoxys Calcitrans (Diptera) Attacking Horses and Cows in Northern Scania, Sweden. *Journal of Biology and Life Science*, *7*(1), 19. https://doi.org/10.5296/jbls.v7i1.8611

Wägele, J. W. (1995). On the information content of characters in comparative morphology and molecular systematics. In *Journal of Zoological Systematics and Evolutionary Research* (Vol. 33, Issues 3–4, pp. 42–47). https://doi.org/10.1111/j.1439-0469.1995.tb00953.x

Wearn, O. R., & Glover-Kapfer, P. (2019). Snap happy: Camera traps are an effective sampling tool when compared with alternative methods. *Royal Society Open Science*, *6*(3). https://doi.org/10.1098/rsos.181748

Wilson, E. O. (2004). Taxonomy as a fundamental discipline. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1444), 739. https://doi.org/10.1098/RSTB.2003.1440