1 Barcoding Brazilian mammals to monitor biological diversity and threats:

2 trends, perspectives, and knowledge gaps

3 Hernani Fernandes Magalhães de Oliveira^{1,2,3*}; Geraldo Brito Freire-Jr.²; Daiana Cardoso Silva⁴;

4 Vanessa Alves Mata^{5,6}; Fernanda Delborgo Abra^{7,8,9}; Nícholas Ferreira Camargo¹⁰; Larissa

5 Gabriela Araujo Goebel¹¹; Gabriela Rodrigues Longo¹²; Joaquim Manoel Silva⁴; Guarino Rinaldi

6 Colli²; Fabricius Maia Chaves Bicalho Domingos¹

7

8 Affiliations

⁹ ¹ Departamento de Zoologia, Universidade Federal do Paraná – UFPR, Curitiba, Brazil

¹⁰ ² Departamento de Zoologia, Universidade de Brasília – UnB, Brasília, Brazil

¹¹ ³ Knowledge Center for Biodiversity, Belo Horizonte, MG, 31270–901, Brazil.

¹² ⁴ Programa de Pós-graduação em Ecologia e Conservação, Universidade do Estado de Mato

13 Grosso – UNEMAT, Nova Xavantina, Brazil

¹⁴ ⁵ CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório

- 15 Associado, Campus de Vairão, Universidade do Porto, Vairão, Vila do Conde, Portugal
- ¹⁶ ⁶ BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão,
- 17 Vairão, Vila do Conde, Portugal
- ¹⁸ ⁷ Smithsonian National Zoo and Conservation Biology Institute—Center for Conservation and
- 19 Sustainability, Washington, DC, USA
- ²⁰ ⁸ ViaFAUNA Estudos Ambientais, São Paulo, SP, Brazil
- ²¹ ⁹ Instituto Pró-Carnívoros, Atibaia, SP, Brazil
- ²² ¹⁰ Departamento de Ecologia, Universidade de Brasília UnB, Brasília, Brazil

- 23 ¹¹ Programa de Pós-graduação em Ciências Ambientais, Universidade do Estado de Mato Grosso
- 24 UNEMAT, Cáceres, Brazil
- ²⁵ ¹² Programa de Pós-graduação em Ensino de Ciências, Universidade Federal de Mato Grosso do
- 26 Sul UFMS, Campo Grande, Brazil
- 27
- *Corresponding author information. E-mail: <u>oliveiradebioh@gmail.com</u>. Full postal address:
 SQN 209, Bloco E Apto 304. City: Brasília. Country: Brazil. Zip code: 70854-050
- 30
- 31
- 32

Abstract

34 DNA barcoding and environmental DNA (eDNA) represent an important advance for biomonitoring the world's biodiversity and its threats. However, these methods are highly 35 36 dependent on the presence of species sequences on molecular databases. Brazil is one of the largest and most biologically diverse countries in the world. However, many knowledge gaps 37 38 still exist for the description, identification, and monitoring of its mammalian biodiversity using 39 molecular methods. We aimed at unraveling the patterns related to the presence of Brazilian mammal species on molecular databases to improve our understanding on how effectively it 40 41 would be to monitor them using DNA barcoding and environmental DNA, and contribute to the 42 conservation of this flagship zoological group. We found that many gaps exist in molecular databases, with many taxa being poorly represented, particularly from the Amazon biome, the 43 44 order Lagomorpha, and arboreal, gomivorous, near extinct, and illegally traded species. Moreover, our analyses revealed that the year of species description was the most important 45 factor determining the probability of a species to have been sequenced. Primates are the group 46 47 with the highest number of species considered to be priority for sequencing due to their high level of combined threats. We highlight where investments are needed to fill knowledge gaps 48 and increase the representativity of species on molecular databases to enable a better monitoring 49 ability of Brazilian mammals encompassing different traits using DNA barcoding and 50 environmental DNA. 51

52

53 Keywords: biomonitoring; conservation; environmental DNA; Mammalia; Neotropics.

1. Introduction

The fast pace of biodiversity decline worldwide demands effective and rapid 55 biomonitoring tools to evaluate species and population trends. However, most traditional 56 methods used in the inventory of terrestrial biodiversity are usually time-consuming either due to 57 58 the high effort to set the equipments in the field and capture the specimens or due to the time to 59 properly process the samples and identify species (Carl et al. 2020; Gogarten, 2020; Mena et al. 2021). Moreover, they can also be inaccurate in terms of species identification due to taxonomic 60 limitations related to morphologically identifying some taxa (Meek et al. 2013). These can be 61 62 particularly problematic for taxa and regions with high levels of cryptic species, such as the 63 Neotropical region (Clare et al. 2007; Domingos et al. 2014; Funk et al. 2012).

64 Morphological identification of mammals can be complex due to the need to evaluate 65 internal anatomical structures (such as skulls or teeth) (Galimberti et al. 2015), and requires specialized taxonomic training and practice, and a comparative reference material in zoological 66 collections, which is often lacking (Francis et al. 2010; Pečnikar and Buzan, 2014), in particular 67 68 in vast unexplored regions of the Neotropics. However, molecular tools developed in the last 69 decades considerably increased the ability to overcome the need to evaluate morphological 70 characteristics for species identification and description (Schindel et al. 2005; Galimbarti et al. 71 2015). The application of DNA barcoding to ecological studies have largely advanced in the last two decades, enabling the genetic identification of mammal species from tissue samples and 72 73 DNA traces recovered from the environment (eDNA), greatly improving species detectability (von Cräutlein et al. 2011; Ivanova et al. 2012; Lynggaard et al. 2023; Saranholi et al. 2023). 74 Moreover, it has also increased our ability to understand the impacts of anthropogenic pressures 75 76 on threatened species (Klippel et al. 2015; Guimarães-Costa et al. 2020). The application of eDNA can yield comprehensive, cost-effective, and representative datasets for different
monitoring conditions and target species (Ji et al. 2013; Mena et al. 2021).

79 The International Barcode of Life (IBOL) initiative is considered to be one of the largest biodiversity consortiums in the world, with a total of five billion dollars invested over the course 80 81 of ten years (Lewin et al. 2018). IBOL aims at sequencing all known multicellular species in the 82 world until 2030 (Lewin et al. 2018; Lewin et al. 2022), which will make it possible to monitor Earth's biodiversity on an unprecedented scale. However, we are still far from achieving this goal 83 for some species groups and countries, hindering our ability to properly monitor the world's biota 84 85 ecological trends and dynamics. Reliable species identification based on DNA relies on the presence of the species sequence on molecular databases (Kvist, 2013; Theissinger et al. 2023). 86 87 Thus, taxonomic groups that are poorly represented in molecular databases are generally harder 88 or impossible to be identified using molecular tools (Kvist, 2013; Theissinger et al. 2023), making it also challenging for these taxa to be monitored and protected. Thus, identifying and 89 filling biodiversity gaps in molecular databases is crucial to properly monitor and protect the 90 91 world's biodiversity (Šmíd, 2022). Countries in tropical regions harbour a significant number of non-sequenced vertebrate species (Šmíd, 2022), due to both challenges in collecting and 92 sequencing tissue samples, and the high biodiversity levels of the regions, making tropical 93 94 species more vulnerable to anthropogenic threats.

In addition to enhancing biodiversity monitoring, DNA barcoding can also improve species identification for different threats, such as poaching, roadkill, and illegal wildlife trade (Sanches et al. 2011; Klippel et al. 2015; Formentão et al. 2021), which are important threats to the conservation of mammals (Costa et al. 2005; Ferreira, 2014; Pinto et al. 2022). One of the main challenges to monitor the impact of these threats is that they often leave unidentifiable

traces of biological tissue for morphological identification (Santos et al. 2012; Klippel et al.
2015), including meat, nails or hair, making it difficult or even impossible for species
morphological detection and identification (Huffman and Wallace, 2012; Johnson et al., 2014).
However, animal remains can be used in the detection of particular species with the use of
environmental DNA, including airDNA or surface swabs (Bohmann et al. 2014; Clare et al.
2022; Drescher et al. 2022; Lynggaard et al. 2023; Gonçalves et al. 2015).

106 Brazil is the fifth largest country in the world and the largest tropical country, representing 107 almost half of the entire area of the Neotropical region (44%), which is one of the most diverse 108 biogeographic regions worldwide (Raven et al. 2020). Furthermore, the country hosts seven 109 biomes, with two biodiversity hotspots (Cerrado and Atlantic Rainforest), as well as the largest 110 and most diverse world's rainforest, the Amazon (Myers et al. 2000; Kintish, 2007). Brazil is 111 also a megadiverse country, with one of the richest mammalian fauna worldwide (775 mammal species) (Mittermeier et al. 1997, Costa et al. 2005; Abreu et al. 2022), representing over 10% of 112 the mammal species on the planet (Mammal Diversity Database, 2023). Similar to other 113 114 megadiverse countries, many cryptic species and previously unknown taxa still remain to be 115 described in Brazil (Pinto et al. 2015; Domingos et al. 2017). On average, more than four new 116 Brazilian mammal species are described every year (Paglia at el. 2012), including recent descriptions of species with a wide range of body sizes, such as bats, primates, and river dolphins 117 118 (Hrbek et al. 2014; Moratell and Dias, 2015; Boubli et al. 2019). Thus, Brazilian high cryptic 119 mammalian diversity and species richness, makes it difficult to taxonomically identify, describe, 120 and monitor mammalian biodiversity within the country, challenging the conservation of this 121 ecologically diverse group.

122 Our study aimed at unraveling the current ability to monitor Brazilian mammals using 123 DNA barcoding and environmental DNA (eDNA) based methods to improve our understanding 124 on how effectively it is possible to monitor biological trends and dynamics. More specifically, 125 our goals were threefold: 1) describe mammal species coverage on molecular databases based on 126 different ecological and conservation traits (distribution, conservation status, threat, order, diet, 127 habit); 2) understand what are the most important traits that predict the probability of a species to 128 be available at molecular databases; 3) create a priority list of threatened Brazilian mammal 129 species to be sequenced for DNA barcoding.

130

131 2. Methods

132

2.1. Dataset about Brazilian mammals

133 Using a list of mammal species compiled by the Brazilian Mammalogical Society (Abreu et al. 2022), we searched for the presence of available Cytochrome c oxidase subunit I (COI) 134 DNA barcode sequences for each species on the main molecular database responsible for storing 135 136 DNA barcodes (Bold Systems), and then categorized each species as being present or absent 137 from the database. Thus, for each Brazilian mammal species, we compiled a set of traits using data from multiple references: species description year, biome of occurrence, diet, habitat, 138 endemism to Brazil, restricted distribution (occurring in an area of less than 20,000 km²), weight, 139 140 IUCN conservation status, population trend, and threats (poaching, roadkill, and illegal wildlife 141 trade) (Supplemental files).

We initially compiled data about biome, diet, habitat, endemic to Brazil, restricted distribution, and weight, using the information available at Paglia et al. (2012). For species that were only described or recognized to occur in Brazil after the publication of Paglia et al. (2012), we complemented the information present on the database with additional literature searches and consults to specialists on different mammal orders. For information about mammal species conservation status and population trends we used IUCN (2021). To assess whether species are known to be affected by roadkills and poaching, we used Grilo et al. (2018) and Praill et al. (2022), and Ferreira (2014), respectively. Finally, we used the CITES database for the list of mammal species that are threatened due to illegal wildlife trade (CITES, 2023).

151

152 *2.2. Statistical analysis*

153 To spatially represent species coverage on molecular databases, we considered the 154 distribution maps of 663 Brazilian mammal species that were available at Map of Life (https://mol.org/species - Marsh et al. (2022)). We used grid cells of 50 km across the whole 155 156 extension of Brazil and plotted the density of mammal richness using the density analysis plugin available at QGIS (https://gis.org). We calculated density following two steps: initially, we 157 considered the 663 mammal species as the total species richness; afterwards, we focused on the 158 159 427 mammal species for which there were DNA sequences available. From this, we calculated 160 the relative density of species with DNA sequences available. We further repeated this analysis 161 for each individual mammal order (Figure S1).

To infer the influence of mammal species biological traits and other important historical, ecological and conservation-related predictors on the probability of a species to have a DNA barcode deposited in public database, we applied a classical machine-learning approach using Classification and Regression Trees (CART, Breiman et al, 1984) implemented in the package RPART (Therneau et al, 2022). The advantages of this algorithm over other machine-learning or linear statistics alternatives is the straightforward use of both continuous and discrete data, and the fact that it can be implemented when there are missing values (NAs) for some predictors without incurring in the complete removal of the case (i.e., mammal species, in our dataset) as would be necessary in linear regression approaches. As usually done for regression or classification machine learning approaches, we partitioned the data into two sets, one for training (70% of the cases) and another for testing (30% of the cases). This approach of training the algorithm with one set of data and testing it with an independent set is a powerful approach that allows for straightforward error estimation.

Hereinafter, we use the term "bold status" in reference to the data of species that either 175 176 have (bold status = yes) or do not have (bold status = no) a COI DNA barcode deposited in the 177 BOLD systems. We run two independent CART analyses, always using bold status as the dependent variable: 1) using all compiled variables as independent variables; 2) excluding all 178 179 predictors for which the data reflects Brazilian-specific information, namely Biome distribution, poaching, roadkill, and Brazilian endemism status (Figure S2 and Figure S3). After both 180 analyses, we calculated the accuracy of each classification tree based on the number of correctly 181 182 identified bold status in the testing set.

183 For the prioritization of the threatened species to be sequenced, we used the following 184 species traits: endemism to Brazil, restricted distribution (distribution within an area of less than 20,000 km²), conservation status (vulnerable, endangered or critically endangered), population 185 186 trend (decreasing), and threats (poaching, hunting, and illegal wildlife trade). Species that 187 matched five of these criteria were listed as extremely high priority, while species matching three 188 to four were listed as high priority, and one to two were considered as intermediate priority. Furthermore, non-threatened species according to IUCN were not included in our list and were 189 190 considered to be low priority for sequencing in relation to conservation purposes.

192 **3. Results**

193

3.1. Species spatial coverage on molecular databases

194 In total, 482 Brazilian mammal species were present on the BOLD system database, 195 which represents 62.19% of Brazilian mammals (775 spp.). The Amazon Forest was the biome with the highest number of species (453 spp.), but also with the highest number of missing 196 197 species on BOLD (181 spp. - 39.96%). On the opposite, the Pampas biome had the lowest number of species (83 spp.), but with the highest proportion of species present on BOLD (71 198 199 spp. - 85.54%). In relation to the distribution of species coverage across the country, the northern 200 and southern portions of Brazil, located in the Amazon Forest and the Pampas had the lowest and 201 the highest coverage, respectively (Figure 1). The western part of Brazil, including mostly the 202 Amazon Forest, had the lowest species coverage, reaching values as low as 75%. Moreover, regarding the representativity of each mammal order, only Chiroptera and Cetartiodactyla had 203 204 higher representativity of species on BOLD in western Amazon, but lower in the Pampas.

205

3.2. Species taxonomic, ecological and conservation representation in molecular
databases

In relation to the Brazilian mammal orders with at least 10 species, Cetartiodactyla (60 spp.) had the highest species coverage on BOLD (47 spp. - 78.33%) and Primates had the lowest (131 spp. - 33.59%). Considering Brazilian mammal diets, insectivorous species were the most common across Brazilian mammals (288 spp. - 37.35%), planctophagous species had the highest species coverage (10 spp. - 90.00%), while gomivorous species had the lowest (22 spp. -18.18%). Regarding species habits, arboreal habit was the most common across Brazilian mammals (240 spp - 31.96%) and also had the lowest species coverage (111 spp. - 46.25%), with
aquatic species having the highest species coverage (46 spp. - 79.31%). In relation to the level of
threat, most Brazilian mammal species were listed as least concern (471 spp.), which was also
the group with the highest species coverage (356 spp. - 75.58%), while not evaluated (15 spp. 15.79%) and critically endangered species (4 spp. - 28.57%) had the lowest coverage.

- 219
- 220

3.3. Traits influencing species DNA barcoding

221 In our first analysis, using all predictors, the species year of description and whether it 222 was listed as least concern according to the IUCN were the most important variables influencing 223 the decision tree algorithm and explaining bold status (Fig. S2). These analyses had an accuracy 224 of 76.7% (i.e., correctly identified bold status) based on the testing set, which is in line with 225 other studies using machine-learning approaches to classify biological data (Domingos et al. 226 2014, Scalon et al. 2020). In our second analysis, excluding Brazilian-specific information, we 227 found that species not evaluated by the IUCN and year of description were the most important 228 variables influencing the decision tree algorithm and explaining bold status (Fig. S2). These analyses had an accuracy of 70.9% (i.e., correctly identified bold status) based on the testing set. 229

230

231

3.4. Representativity of species under threat

In relation to some of the main known threats for Brazilian mammals, roadkill (93 spp.) was the one with the highest number of species coverage on molecular databases (84.54%); followed by poaching (123 spp.) (84 spp. - 68.29%), and illegal wildlife trade (187 spp.) with the lowest species coverage (60.43%). Regarding the species coverage for these threats in relation to each mammal order (with more than one species), Pilosa, Cingulata and Chiroptera were the ones with the highest species coverage (100% for all threats), while Cetarrtiodactyla, Primates,
Rodentia, and Sirenia showed the lowest species coverage (Figure 4).

- 239
- 240

3.5. Prioritization for species DNA barcoding

241 Forty five Brazilian mammal species are currently listed as threatened according to the 242 IUCN Red list, but still do not have publicly available DNA barcodes according to our results. 243 From those, we listed seven species as extremely high priority, 35 species as high priority and 244 three as moderate priority for DNA barcoding (Table S1). The most important species to be 245 sequenced (extremely high priority) due to their high level of threat coupled with their endemism 246 to Brazil and restricted distribution were: Marca's marmoset (Mico marcai - Primates), Rondon's 247 mamoset (*Mico rondoni* - Primates), Maranhão red-handed howler (*Alouatta ululata* - Primates), 248 Coimbra Filho's titi monkey (Callicebus coimbrai - Primates), black-faced lion tamarin (Leontopithecus caissara - Primates), buffy-headed marmoset (Callithrix flaviceps - Primates), 249 250 and pied tamarin (Saguinus bicolor - Primates).

251

4. Discussion

Many gaps still exist in the ability to identify and monitor Brazilian mammals and their threats using DNA barcoding and environmental DNA, considering the available COI data in public molecular databases. The Amazon biome (more specifically its western region), the order Lagomorpha, arboreal, gomivorous, near extinct, and illegally traded species, are proportionally the taxa with the lowest representativity in molecular databases. Furthermore, the most important traits that define the likelihood of a taxa to have been sequenced are the year of description and its conservation status on IUCN, with long-described species and of least concern having a higher probability of being represented in DNA barcode libraries than others. Lastly, primates
were the most speciose threatened taxa that needs to be sequenced to increase species coverage
in molecular databases.

- 263
- 264

4.1. Species spatial coverage on molecular databases

Our results follow the general finding that 81% of mammal species worldwide have genetic data available on molecular databases (Šmíd, 2022). The differences in species coverage between biomes might be related with an effect of the number of species with a restricted distribution (Šmíd, 2022), which would restrict certain species to one or few biomes, making it also harder to acquire biological samples for sequencing. Furthermore, biomes with lower species richness, such as the Pampas, require lower sampling and sequencing efforts to achieve a higher species coverage in molecular databases.

The Amazon biome was not only the biome with the poorest species representation on 272 273 molecular databases, but also held the grid cells with the lowest species richness coverage across 274 Brazil. These could be related with the low accessibility of the Amazon, especially in its western portion, which has limited access from roads (Grilo et al. 2021). Our results are partially 275 congruent with Šmíd (2022), which pointed out that the Northern Amazon is one of the best 276 277 represented portions of the world for mammals, but are discordant with the results in comparison with the western Amazon, which was one of the poorest represented areas in our study. This 278 279 might be related with an increase in the availability of distribution maps for mammals and newly 280 described species (Marsh et al. 2022; Rengifo et al. 2022), which revealed more specific regional patterns for species that were not considered in Šmíd (2022). 281

Chiroptera and Cetartiodactyla were relatively well represented in BOLD across the western Amazon, which is the opposite trend compared to other mammalian orders, and might be explained by species of both these orders having large distributions (Šmíd, 2022). Many species within both orders have wide distributions covering large proportions of the Neotropical region (Marsh et al. 2022), which would make it easier for biological samples of these species to be collected and sequenced in other more accessible biomes or surrounding countries.

288

289 4.2. Species taxonomic, ecological and conservation representation in molecular
290 databases

291 Mammals are a highly diverse order, which includes species that have many different diets, habits, and levels of threat (Paglia et al. 2012). In order for DNA barcoding to be an 292 293 effective tool for the detection of species with these characteristics, these species need to be 294 sequenced and deposited in molecular databases (Balint et al. 2018; Blaxter et al. 2022). 295 However, the molecular database that we evaluated is highly skewed towards species groups or 296 with specific traits (Cetartiodactyla, planctophagous, aquatic, and least concern species), which hinders our ability to monitor taxa with different traits in the landscape as well as fully 297 understand ecosystem dynamics. 298

These molecular availability trends are probably related to the characteristics of the species from each group. Arboreal species, for example, represent close to a third of Brazilian mammal species, but have less than half of the total species present on molecular databases. This could be related with the fact that canopies still remain an ecological frontier and one of the least accessible places on land (Erwin, 1983; Nakamura et al. 2017). Moreover, many arboreal mammals tend to be elusive and rarely come down to the ground (Eppley et al. 2022; Kaizer etal. 2022), increasing the challenges to collect biological samples for further sequencing.

306 It is alarming for species conservation that there are still large gaps for threatened 307 Brazilian mammals on molecular databases, with their representativity ranging from 25% to 308 60%. The poor representation of these groups in relation to less threatened species can also be 309 related with shared traits among threatened species, such as small populations and geographic 310 ranges, which also makes it difficult for biological samples from these species to be collected for sequencing. Furthermore, it is interesting to note that species that were not evaluated in relation 311 312 to their conservation status had the worst representativity ($\sim 10\%$), which might be further 313 leading to a low interest in studying them, and consequently creating a gap in their representation 314 on molecular databases. Alternatively, their lack of evaluation can also be related with species 315 that were only recently described and, thus, still do not have biological samples available for sequencing. 316

317

318

4.3. Traits influencing species sequencing

319 Our CART results provide clear evidence that researchers were better able to obtain and 320 sequence samples from non-threatened species, and those who have been described for a long 321 time. We interpret this finding in two non-exclusive ways: over the years, it was easier for 322 researchers to obtain and sequence samples from species that are not in some important threat 323 category (more abundant and easier to obtain biological samples), and species that have been 324 taxonomically available for more time were also more prone to be sequenced. Moreover, it is important to highligh that most Brazilian mammal species are not threatened (504 spp - 65.03%), 325 326 which is probably associated with the fact that most species have not been evaluated by the 327 IUCN. Although it is reasonable to hypothesize that some biological traits such as having a 328 widespread distribution and displaying larger sizes would influence the availability of specimens 329 for obtaining samples and sequencing, these had minor influences on the decision tree in relation 330 to the best predictors (Fig. S3). Nonetheless, supporting this hypothesis, the biological traits that 331 were important in explaining bold status were weight, and the distribution in the Atlantic Forest 332 and Marine biomes, indicating that these characteristics indeed influenced the availability of specimens for researchers. In this regard, it is important to highlight that all major Brazilian 333 universities and research institutions are located in richer states within the Atlantic Forest biome 334 335 and in the southeast coast, mostly in the states of São Paulo and Rio de Janeiro.

336 The second analysis also reinforces the fact that year of description is an important 337 predictor of bold status, which could be related with researchers having additional time to get 338 biological samples and sequence earlier described species. Furthermore, the biological traits weight, arboreal locomotion and gomivory were important predictors. Taken together, our 339 CART results support the idea that focus should be given to sequence more recently described 340 341 species. Also, as would be expected, at least one biological trait related to species body size (i.e., weight) is important in terms of researchers gaining access to samples, highlighting the need to 342 invest in smaller species. 343

344

345

4.4. Molecular identification of species under threat

Road mortality, poaching, and illegal wildlife trade are some of the main direct threats to the Brazilian mammalian fauna (Galletti et al. 2021). We found that roadkills and poaching are the threats more likely to be monitored using DNA barcoding, but illegal wildlife trade still remains a challenge, with only 60% of species coverage of threatened species. Additionally, primates and rodents are groups that need increased efforts for sequencing threatened species, asthey are still largely under-represented on BOLD.

352 Brazil has the fourth largest road network worldwide (DNIT, 2019), and it is essential that a proper biomonitoring scheme is implemented to measure animal-vehicle collisions. In São 353 Paulo State alone, one out of 26 Brazilian states, it has been estimated that around 40,000 354 355 medium and large wild mammals are roadkilled each year, including a high number of casualties of endangered species (Abra et al. 2021). Thus, roadkill potentially causes significant impacts on 356 mammal populations and even local species extinctions (Grilo et al. 2021), demanding further 357 358 investigations, where DNA barcoding can give a significant contribution. DNA barcoding is of 359 utmost importance in situations where roadkill monitoring is conducted by non-specialists, like 360 road maintenance personnel on private toll roads in Brazil, for instance. A study conducted by 361 Abra et al. (2018) revealed that non-specialists lack the capacity to identify rare or infrequently observed species, those that closely resemble other species (such as small wild canids and 362 363 felids), or species that are not easily distinguishable. Consequently, these individuals often 364 misidentify, provide ambiguous descriptions, or fail to identify these species at all, resulting in 365 an unreliable roadkill database. One additional aspect that needs to be addressed is that 366 biological samples of fresh roadkilled animals can also be sequenced for high-quality reference genomes (Allio et al. 2021), which could help to fill molecular databases and help monitoring 367 this threat. 368

Furthermore, Brazilian mammals can be heavily impacted by poaching, with over 15% of mammal species (123 spp.) impacted by this threat, which represents a significant proportion of the total number of species threatened by poaching worldwide (301 spp) (Ripple et al. 2016). Moreover, poaching can lead to severe mammal population declines or even local species

373 extinctions (Jorge et al. 2013; Galetti et al. 2016). However, it is a difficult threat to monitor as 374 several species are reduced to unidentifiable parts after being hunted (remains of bones, organs, 375 hair, meat and skin), making it often almost impossible to morphologically identify them. In 376 addition, poaching is a crime according to the Brazilian environmental law since 1967, and 377 suspects may face detention and financial penalties, but that needs evidence for legal actions to 378 be properly conducted. Thus, DNA barcoding can significantly help to fight this threat and legally prosecute poachers, as it can allow species identification from biological remains. 379 Moreover, if an endangered species is proven to have been killed, the punishment can be more 380 381 severe under Brazilian laws (Di Rocco and Anello, 2021).

382 Finally, illegal wildlife trade can be an important source of introduction not only of 383 invasive mammal species, but also different pathogens (Rosa et al. 2018), and exert important 384 pressures that might lead to populations extinction (Scheffers et al. 2019), which requires an intensive biomonitoring program. Tackling illegal wildlife trade, which threatens around one 385 fourth of mammals in Brazil and globally, demands the identification of where and which 386 387 species are being traded (Scheffers et al. 2019). However, this is the threat where Brazilian mammals are less represented on molecular databases (~60%), limiting actions to fight against it. 388 389 Thus, more investments are needed to increase the representativity of Brazilian mammal species 390 threatened due to illegal wildlife trade on molecular databases. Particular attention needs to be given for species inhabiting the Atlantic Rainforest, the Amazon Forest, and the Cerrado, as 391 392 these biomes account for hotspots of illegally traded mammal species (Scheffers et al. 2019). 393 Moreover, since traded wildlife is often transported on hidden compartments, which are difficult 394 to detect and monitor, environmental DNA from the air and swabs coupled with portable DNA 395 sequencers (e.g., MinION, Oxford Nanopore Technologies) could help tackling this threat in a

more efficient way (Clare et al. 2022; Lynggaard et al. 2023). This is particularly important
because, even though Neotropical mammals are mostly traded as pets, trading for products, such
as clothing, medicines or religious purposes, can also represent an important issue, especially in
the western Amazon (Scheffers et al. 2019).

- 400
- 401

4.5. Prioritization of species sequencing

402 Primates represent the third most speciose mammal order in Brazil, representing almost 403 one sixth of the total Brazilian mammal species richness (16.90%). Nonetheless, over two thirds 404 of priority Brazilian mammal species to be sequenced are primates (31 spp. - 68.89%). Many 405 neotropical primate species not only have a small distribution (Marsh et al. 2022), but are also 406 elusive due to their arboreal habits, which makes them more susceptible to extinction due to 407 habitat loss, and also harder to collect biological samples for sequencing. Moreover, the Amazon 408 is the most speciose Brazilian biome for primates (Paglia et al. 2012), but many places are hard 409 to access due to the lack of roads or other types of access through parts of intact vegetation 410 (Grilo et al. 2021). Primates are highly susceptible to deforestation as many species cannot move 411 very far on a deforested landscape or survive in the absence of trees (Boyle and Smith, 2010; 412 Mendes-Oliveira et al. 2017). Altogether, these factors are likely to explain the dominance of 413 primates as high priority for sequencing in Brazil.

Sequencing priority species is an important task to be conducted at a fast pace, as it will enable them to be properly detected and their populations monitored across the landscape (Monchamp et al. 2023). Moreover, it can make a significant difference in terms of increasing endangered species detection to combat poaching and wildlife trade, but also improving species identification after roadkills, which can help in elaborating strategies for preventing these accidents. This would be an important step to promote the conservation of the species on this list,
but it is also necessary to promote the use of DNA sequences to monitor Brazilian wildlife across
the country. This methodology is still not widely used to monitor wildlife in Brazil due to its
high costs and lack of sequencing facilities in some parts of the country.

- 423
- 424

4.6. Perspectives and challenges for sequencing Brazilian mammals

425 Around 3,000,000 dollars have been invested by the Brazilian Barcode of Life initiative for sequencing Brazilian biodiversity during the period of 2010-2014, which involved over 100 426 427 participating groups and 500 people from different institutions around the country (Eizirik, 428 2015). This initiative has initially led to the generation of barcodes from 2,122 specimens belonging to 344 mammal species, which represents ~45% of the current number of mammal 429 430 species in Brazil (Eizirik, 2015; Abreu et al. 2022). However, progress towards achieving a comprehensive understanding of Brazilian mammal genetic diversity is still far from complete, 431 432 with less than 65% of Brazilian mammal species currently barcoded. Clearly, more investments 433 are needed to speed up this process and achieve the goals of the Earth Biogenome project, a 434 difficulty shared with other similar initiatives, and in line with the results they have reached so 435 far (Threlfall and Blaxter, 2021; Darwin Tree of Life Project Consortium, 2022). The new 436 barcoding phase may be viewed as more challenging since most of the target species remaining 437 are primarily found in the Amazon biome, and the majority exhibit arboreal behavior, making 438 them less susceptible to being roadkilled, poached, or trafficked.

Even though Brazil generally has a better infrastructure for processing biological samples and sequencing species in comparison to other countries in the Global South (Colella et al. 2023), this infrastructure is still lacking in some regions. In remote areas of Brazil, such as the 442 western Amazon, where there is a lack of access to roads, electricity, and research facilities, 443 sequencing technologies that are less dependent on electricity and internet access, and also lighter and more mobile are needed (Ebenezer et al. 2022). Thus, some parts of the country still 444 445 need more investments for building sequencing facilities for biodiversity description and 446 monitoring. This has also led to large knowledge gaps regarding the presence of some taxonomic 447 groups on molecular databases (Eizirik, 2015). Furthermore, sequencing costs remain comparatively high in Brazil after taking into account the conversion to the Brazilian currency, 448 which hinders large country efforts to locally sequence Brazilian species at a fast pace. 449 450 Moreover, for biological samples that were collected in Brazil but sequenced elsewhere, there is 451 a need to make an equal share of the benefits according to the Nagoya Protocol (Colella et al. 452 2023). This is an important part of the discussion on sequencing tropical biodiversity in relation 453 to the debate about the agreement of the Kunming-Montreal Global Biodiversity Framework (KM-GBF) (Hughes, 2023). Thus, it is important that at least part of the benefits from Digital 454 Sequencing Information (DSI) that were sequenced outside the tropics, but which the biological 455 456 samples were originally collected in tropical countries, is reinvested in the countries of origin to 457 increase their infrastructure and capacity building.

There are still many challenges for DNA barcoding to be properly used to describe new species and on mammal biodiversity monitoring programs in developing countries, such as the lack of financial investments, capacity building, and infrastructure (Eizirik, 2015; Paknia et al., 2015). The lack of taxonomists, the high prices of sequencing in the country, the lack of infrastructure together with its large territorial area and high mammalian cryptic diversity are among the main challenges that limit the advance of using DNA barcoding to study Brazilian mammal diversity. Furthermore, Brazilian researchers face a highly beaurocratic process in order to export biological samples to be sequenced outside the country (Colella et al. 2023). However,
advancing our knowledge regarding these gaps in megadiverse countries will significantly
increase our capacity to monitor biodiversity and fill the Linnean, Wallacean and Darwinian
shortfalls in some of the most biodiversity rich parts of the planet (Hortal et al. 2015; Šmíd,
2022; Vergara-Asenjo et al. 2023).

470 Finally, one last challenge that needs to be addressed is that the presence of a species 471 sequence on a molecular database does not fully guarantee its future identification, because of the high genetic divergence of some populations (Gaubert et al. 2014; Galimberti et al. 2015). In 472 473 these cases, sequences from multiple biological samples spanning different geographic locations 474 are needed to properly identify these species (Gaubert et al. 2014; Galimberti et al. 2015). Thus, more efforts are needed to evaluate the genetic variability representation of these species on 475 476 molecular databases, and additional targeted sampling efforts to increase genetic variability on molecular databases. 477

478

479 **5.** Conclusion

480 Even though Brazil is a large and highly species diverse country that requires a high 481 investment to enable sequencing its biodiversity, large advances have been made for Brazilian mammal species in the last 15 years, which enabled a significant increase in monitoring capacity 482 using DNA barcoding and environmental DNA. However, many sequencing gaps still exist, 483 484 especially in relation to specific biomes, traits, diets, taxonomic groups, and threats. Even though 485 time since species description has been the most important factor to determine the sequencing probability of a Brazilian mammal species, now that the gaps for species in molecular databases 486 487 have been identified, more targeted efforts are needed to increase species representativity and

genetic variability to enable the use of DNA techniques to sample and monitor Brazilianmammal biodiversity.

Given that DNA barcoding is a standardized, reliable and precise method, this should facilitate biodiversity assessment, enhance species identification, especially detection of cryptic species (e.g. arboreal mammals), wildlife forensics, conservation genetics and the conservation of endangered species. These efforts will also help building important prospects for the study and understanding of Neotropical mammals ecology and evolution.

495	Competing interests statement
496	The authors have no competing interests to declare in the publication of the current study.
497	
498	
499	Declaration of Generative AI and AI-assisted technologies in the writing process
500	The authors would like to disclose that they have not used any form of artificial
501	intelligence for writing the current article.
502	
503	Authors contributions
504	The roles are: Conceptualization: HFMO; GBF; JMS and FMCBD; Data curation: HFMO;
505	GBFreire-Jr.; VAM; FA; NFC; LGAG; GRL; Formal analysis: DCS, FMCBD; Funding
506	acquisition: HFMO; GRC; FMCBD; Investigation: HFMO; Methodology: HFMO and FMCBD;
507	Project administration: HFMO and FMCBD; Supervision: FMCBD; Roles/Writing - original
508	draft: HFMO; and Writing - review & editing: all authors.
509	
510	Acknowledgements
511	Hernani F. M. Oliveira would like to thank the Coordenação de Aperfeçoamento de Ensino
512	Superior (CAPES) for the funding provided for his postdoctoral scholarship to conduct this
513	project (Processo n 88887.890117/2023-00).
514	
515	Funding: This work was supported by the Coordenação de Aperfeçoamento de Ensino
516	Superior (CAPES) (Grant number: 88887.890117/2023-00).
517	

518 Data statement

- 519 By the time the article is accepted, the dataset regarding this study will be made available
- 520 at: <u>https://github.com/Hernanif/Data_accessibility-</u>
- 521 <u>Barcoding_mammals/tree/10e0b215055f2af27091338e7141dfb3246cb81f</u>

522 References

- Abra, F. D.; Huijser, M. P.; Magioli, M.; Bovo, A. A. A. & de Barros, K. M. P. M. (2021). An
 estimate of wild mammal roadkill in São Paulo state, Brazil. *Heliyon*, 7(1): 1-12.
 https://doi.org/10.1016/j.heliyon.2021.e06015
- Abra, F. D.; Huijser, M. P.; Pereira, C. S. & Ferraz, K. M. (2018). How reliable are your data?
 Verifying species identification of road-killed mammals recorded by road maintenance
 personnel in São Paulo State, Brazil. *Biological Conservation*, 225: 42-52.
 https://doi.org/10.1016/j.biocon.2018.06.019
- 530 [dataset] Abreu, E. F; Casali, D.; Costa-Araújo, R.; Garbino, G. S. T.; Libardi, G. S.; Loretto, D.;
- 531 Loss, A. C.; ... & Tirelli FP. (2022). Lista de Mamíferos do Brasil, v2022-1. Zenodo.
 532 https://doi.org/10.5281/zenodo.7469767
- Allio, R.; Tilak, M. K.; Scornavacca, C.; Avenant, N. L.; Kitchener, A. C.; Corre, E.; ... &
 Delsuc, F. (2021). High-quality carnivoran genomes from roadkill samples enable
 comparative species delineation in aardwolf and bat-eared fox. *Elife*, 10: e63167.
 https://doi.org/10.7554/eLife.63167
- Balint, M.; Pfenninger, M.; Grossart, H. P.; Taberlet, P.; Vellend, M.; Leibold, M. A.; ... &
 Bowler, D. (2018). Environmental DNA time series in ecology. *Trends in Ecology & Evolution*, 33(12): 945-957. https://doi.org/10.1016/j.tree.2018.09.003
- 540 Blaxter, M.; Archibald, J. M.; Childers, A. K.; Coddington, J. A.; Crandall, K. A.; Di Palma, F.;
- 541 ... & Lewin, H. A. (2022). Why sequence all eukaryotes?. Proceedings of the National
- 542 *Academy of Sciences*, 119(4): e2115636118. https://doi.org/10.1073/pnas.2115636118

Bohmann, K.; Evans, A.; Gilbert, M. T. P.; Carvalho, G. R.; Creer, S.; Knapp, M.; ... & De
Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29(6): 358-367. https://doi.org/10.1016/j.tree.2014.04.003

546 Boubli, J. P.; Byrne, H.; da Silva, M. N.; Silva-Júnior, J.; Araújo, R. C.; Bertuol, F.; ... & Hrbek,

- 547 T. (2019). On a new species of titi monkey (Primates: Plecturocebus Byrne et al., 2016),
- from Alta Floresta, southern Amazon, Brazil. *Molecular Phylogenetics and Evolution*, 132:
- 549 117-137. https://doi.org/10.1016/j.ympev.2018.11.012
- Breiman, L.; Friedman, J. H.; Olshen, R. A. & Stone, C. J. (1984). Classification and Regression
 Trees. Chapman and Hall, first ed. Longman, New York.
- Boyle, S. A. & Smith A. T. (2010). Can landscape and species characteristics predict primate
 presence in forest fragments in the Brazilian Amazon? *Biological Conservation*, 143:
 1134-1143. https://doi.org/10.1016/j.biocon.2010.02.008
- 555 Carl, C.; Schönfeld, F.; Profft, I.; Klamm, A. & Landgraf, D. (2020). Automated detection of
- 556 European wild mammal species in camera trap images with an existing and pre-trained
- 557 computer vision model. *European Journal of Wildlife Research*, 66: 1-7.
 558 https://doi.org/10.1007/s10344-020-01404-y
- Clare, E. L.; Lim, B. K.; Engstrom, M. D.; Eger, J. L. & Hebert, P. D. (2007). DNA barcoding of
 Neotropical bats: species identification and discovery within Guyana. *Molecular Ecology Notes*, 7(2): 184-190. https://doi.org/10.1111/j.1471-8286.2006.01657.x
- 562 Clare, E. L.; Adams, A. M.; Maya-Simões, A. Z.; Eger, J. L.; Hebert, P. D. & Fenton, M. B.
- 563 (2013). Diversification and reproductive isolation: cryptic species in the only New World
- 564 high-duty cycle bat, *Pteronotus parnellii*. *BMC Evolutionary Biology*, 13(1): 1-18.
- 565 https://doi.org/10.1186/1471-2148-13-26

566 Clare, E. L.; Economou, C. K.; Bennett, F. J.; Dyer, C. E.; Adams, K.; McRobie, B.; ... &
567 Littlefair, J. E. (2022). Measuring biodiversity from DNA in the air. *Current Biology*,

568 32(3): 693-700. https://doi.org/10.1016/j.cub.2021.11.064

- 569 CITES, Checklist of CITES species, 2022. Available at: https://checklist.cites.org/. Accessed on
 570 20 June 2023
- 571 Colella, J. P.; Silvestri, L.; Súzan, G.; Weksler, M.; Cook, J. A. & Lessa, E. P. (2023). Engaging
 572 with the Nagoya Protocol on Access and Benefit-Sharing: recommendations for
 573 noncommercial biodiversity researchers. *Journal of Mammalogy*, 104(3): 430-443.
 574 https://doi.org/10.1093/jmammal/gyac122
- 575 Costa, L. P.; Leite, Y. L. R.; Mendes, S. L. & Ditchfield, A. D. (2005). Mammal conservation in
 576 Brazil. *Conservation Biology*, 19(3): 672-679. https://doi.org/10.1111/j.1523577 1739.2005.00666.x
- 578 Darwin Tree of Life Project Consortium. (2022). Sequence locally, think globally: The Darwin 579 tree of life project. *Proceedings of the National Academy of Sciences*, 119(4):
- 580 e2115642118. <u>https://doi.org/10.1073/pnas.211564211</u>
- 581 Di Rocco, F. & Anello, M. (2021). The use of forensic DNA on the conservation of neotropical
 582 mammals. in: Nardelli, M., Túnez, J. I. (Eds.), Molecular Ecology and Conservation
 583 Genetics of Neotropical Mammals. Springer, New York, pp. 85-98.
- 584 DNIT Departamento Nacional de Infraestrutra de Transporte, 2019. Rodovias Federais.
 585 Available at: http://www.dnit.gov.br/. Accessed in: 18/04/202
- 586 Domingos, F. M. C. B.; Bosque, R. J.; Cassimiro, J.; Colli, G. R.; Rodrigues, M. T.; Santos, M.
- 587 G., & Beheregaray, L.B. (2014). Out of the deep: Cryptic speciation in a Neotropical

gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods. *Molecular Phylogenetics and Evolution*, 80(1): 113-124. https://doi.org/10.1016/j.ympev.2014.07.022

590 Domingos, F. M.; Colli, G. R.; Lemmon, A.; Lemmon, E. M. & Beheregaray, L. B. (2017). In

- 591 the shadows: phylogenomics and coalescent species delimitation unveil cryptic diversity in
- a Cerrado endemic lizard (Squamata: Tropidurus). *Molecular Phylogenetics and Evolution*,
- 593 107: 455-465. https://doi.org/10.1016/j.ympev.2016.12.009
- Drescher, L.; Heng, N. J. K.; Chin, M. Y.; Karve, N. R.; Cheung, E. J. W.; Kurniadi, A.; ... &
 Wainwright, B. J. (2022). Blood in the water: DNA barcoding of traded shark fins in
 Singapore. *Frontiers in Marine Science*, 9: 1-7. https://doi.org/10.3389/fmars.2022.907714
- 597 Ebenezer, T. E.; Muigai, A. W.; Nouala, S.; Badaoui, B.; Blaxter, M.; Buddie, A. G.; ... &
 598 Djikeng, A. (2022). Africa: sequence 100,000 species to safeguard biodiversity. *Nature*,
 599 603(7901): 388-392. https://doi.org/10.1038/d41586-022-00712-4
- Eizirik, E. (2015) The Brazilian Barcode of Life (BrBOL) initiative and its potential to aid in
 biodiversity conservation: an overview of the effort to catalogue mammals, reptiles
 (including birds), and amphibians from a megadiverse region. *Genome*, 58(5): 214.
- 603 Eppley, T. M.; Hoeks, S.; Chapman, C. A.; Ganzhorn, J. U.; Hall, K.; Owen, M. A.; ... & Santini,
- 604 L. (2022). Factors influencing terrestriality in primates of the Americas and Madagascar.
- 605 Proceedings of the National Academy of Sciences, 119(42): e2121105119.
- 606 https://doi.org/10.1073/pnas.212110511
- Erwin, T. L. (1983). Tropical forest canopies: the last biotic frontier. *Bulletin of the ESA*,
 29(1):14-20. https://doi.org/10.1093/besa/29.1.14
- Ferreira, H. F. (2014). A caça no Brasil: panorama histórico e atual. PhD thesis Universidade Federal da Paraíba. 466 pp.

- Fišer, P. Ž. & Buzan, E. V. (2014). 20 years since the introduction of DNA barcoding:
 from theory to application. *Journal of Applied Genetics*, 55: 43-52.
 https://doi.org/10.1007/s13353-013-0180-y
- 614 Formentão, L.; Saraiva, A. S. & Marrero, A. R. (2021). DNA barcoding exposes the need
- to control the illegal trade of eggs of non-threatened parrots in Brazil. *Conservation*
- 616 *Genetics Resources*, 13(3): 275-281. https://doi.org/10.1007/s12686-021-01209-4
- 617 Francis, C. M.; Borisenko, A. V.; Ivanova, N. V.; Eger, J. L.; Lim, B. K.; Guillén-Servent,
- A.; ... & Hebert, P. D. N. (2010). The role of DNA barcodes in understanding and
 conservation of mammal diversity in Southeast Asia. *PLoS ONE*, 5(9): e12575.
 https://doi.org/10.1371/journal.pone.0012575
- Funk, W. C.; Caminer, M. & Ron, S. R. (2012). High levels of cryptic species diversity
 uncovered in Amazonian frogs. *Proceedings of the Royal Society B: Biological Sciences*,
 279(1734): 1806-1814. https://doi.org/10.1098/rspb.2011.1653
- 624 Galetti, M.; Brocardo, C. R.; Begotti, R. A.; Hortenci, L.; Rocha-Mendes, F.; Bernardo, C. S. S.;
- 625 ... & Siqueira, T. (2016). Defaunation and biomass collapse of mammals in the largest
 626 Atlantic forest remnant. *Animal Conservation*, 20(3): 270-281.
 627 https://doi.org/10.1111/acv.12311
- 628 Galetti, M.; Gonçalves, F.; Villar, N.; Zipparro, V. B.; Paz, C.; Mendes, C.; ... & Bovendorp, R.
- 629 S. (2021). Causes and consequences of large-scale defaunation in the Atlantic forest. in:
- 630 Marques, M. C. M. and Grelle, C. E. V. (Eds.). The atlantic forest: history, biodiversity,
- threats and opportunities of the mega-diverse forest. Springer, New York, 297-324.

- Galimberti, A.; Sandionigi, A.; Bruno, A.; Bellati, A. & Casiraghi, M. (2015). DNA barcoding in
 mammals: what's new and where next?. *Hystrix*, 26(1): 13-24.
 https://doi.org/10.4404/hystrix-26.1-11347
- 635 Gaubert, P.; Njiokou, F.; Olayemi, A.; Pagani, P.; Dufour, S.; Danquah, E.; ... & Antunes, A.
- 636 (2014). Bushmeat genetics: setting up a reference framework for the DNA typing of
 637 African forest bushmeat. *Molecular Ecology Resources*, 15(3): 633–651.
- 638 https://doi.org/10.1111/1755-0998.12334
- 639 Gogarten, J. F.; Hoffmann, C.; Arandjelovic, M.; Sachse, A.; Merkel, K.; Dieguez, P.; ... &
- 640 Calvignac-Spencer, S. (2020). Fly-derived DNA and camera traps are complementary tools
- 641 for assessing mammalian biodiversity. *Environmental DNA*, 2(1): 63-76.
 642 https://doi.org/10.1002/edn3.46
- 643 Grilo, C.; Coimbra, M. R.; Cerqueira, R. C.; Barbosa, P.; Dornas, R. P; Gonçalves, L. O; ... &
- Kindel, A. (2018). BRAZIL ROAD-KILL: a data set of wildlife terrestrial vertebrate roadkills. *Ecology*, 99(11): 2625-2625. https://doi.org/10.1002/ecy.2464
- 646 Grilo, C.; Borda-de-Água, L.; Beja, P.; Goolsby, E.; Soanes, K.; le Roux, A.; ... & González-
- 647 Suárez, M. (2021). Conservation threats from roadkill in the global road network. *Global*
- 648 *Ecology and Biogeography*, 30(11): 2200-2210. https://doi.org/10.1111/geb.13375
- 649 Gonçalves, P. F.; Oliveira-Marques, A. R.; Matsumoto, T. E. & Miyaki, C. Y. (2015).
- 650 DNA barcoding identifies illegal parrot trade. *Journal of Heredity*, 106(S1): 560-564.
- 651 https://doi.org/10.1093/jhered/esv035
- Guimarães-Costa, A.; Machado, F. S.; Reis-Filho, J. A.; Andrade, M.; Araújo, R. G.;
- 653 Corrêa, E. M. R.; ... & Giarrizzo, T. (2020). DNA Barcoding for the assessment of the
- taxonomy and conservation status of the fish bycatch of the northern Brazilian shrimp

- 655 trawl fishery. *Frontiers in Marine Science*, 7: 566021.
 656 https://doi.org/10.3389/fmars.2020.566021
- Huffman, J. E. & Wallace, J. R. (2012). Wildlife forensics: methods and applications. John
 Wiley & Sons, first ed. Chichester, West Sussex.
- Hughes, A. C. (2023). The Post-2020 Global Biodiversity Framework: How did we get here, and
 where do we go next? *Integrative Conservation*, 2(1): 1-9. https://doi.org/10.1002/inc3.16
- 661 IUCN. (2021). The IUCN Red List of threatened species Version 2021–3. Available at:
 662 http://iucnredlist.org. Accessed on 30 June 2023.
- Johnson, R. N.; Wilson-Wilde, L. & Linacre A. (2014). Current and future directions of DNA in
- wildlife forensic science. Forensic Science International Genetics, 10: 1–11.
 https://doi.org/10.1016/j.fsigen.2013.12.007
- Jorge, M. L. S. P.; Galetti, M.; Ribeiro, M. C. & Ferraz, K. M. P. M. B. (2013). Mammal
 defaunation as surrogate of trophic cascades in a biodiversity hotspot. *Biological Conservation*, 163, 49–57. https://doi.org/10.1016/j.biocon.2013.04.018
- 669 Kaizer, M. C.; Alvim, T. H.; Novaes, C. L.; Mcdevitt, A. D. & Young, R. J. (2022). Snapshot of
- the Atlantic Forest canopy: surveying arboreal mammals in a biodiversity hotspot. *Oryx*,
 56(6): 825-836. http://10.1017/S0030605321001563
- 672 Hortal, J.; de Bello, F.; Diniz-Filho, J. A. F.; Lewinsohn, T. M.; Lobo, J. M. & Ladle, R. J.
- 673 (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of*
- 674 Ecology, Evolution, and Systematics, 46: 523-549. https://doi.org/10.1146/annurev-
- 675 ecolsys-112414-054400

- 676 Hrbek, T.; da Silva, V. M. F.; Dutra, N.; Gravena, W.; Martin, A. R. & Farias, I. P. (2014). A
- 677 new species of river dolphin from Brazil or: how little do we know our biodiversity. *PLoS*678 *ONE*, 9(1): e83623. https://doi.org/10.1371/journal.pone.0083623
- 679 Ivanova, N. V.; Clare, E. L. & Borisenko, A. V. (2012). DNA barcoding in mammals. in: Kress,
- 680 W. J. and Erickson, D. L. (Eds.). DNA barcodes: methods and protocols. Humana Press,
 681 New Jersey, pp. 153-182.
- 682 Ji, Y.; Ashton, L.; Pedley, S. M.; Edwards, D. P.; Tang, Y.; Nakamura, A.; ... & 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
- Kintisch, E. (2007). Improved monitoring of rainforests helps pierce haze of deforestation. –
 Science, 316: 536–537.http://10.1126/science.316.5824.536
- 687 Klippel, A. H.; Oliveira, P. V.; Britto, K. B.; Freire, B. F.; Moreno, M. R.; Dos Santos, A.
- R.; ... & Paneto, G. G. (2015). Using DNA barcodes to identify road-killed animals in two
 atlantic forest nature reserves, Brazil. *PLoS ONE*, 10(8): e0134877.
 https://doi.org/10.1371/journal.pone.0134877
- Kvist, S. (2013). Barcoding in the dark?: a critical view of the sufficiency of zoological DNA
 barcoding databases and a plea for broader integration of taxonomic knowledge. *Molecular Phylogenetics and Evolution*, 69(1): 39-45. https://doi.org/10.1016/j.ympev.2013.05.012
- 694 Lewin, H. A.; Robinson, G. E.; Kress, W. J.; Baker, W. J.; Coddington, J.; Crandall, K. A.; ... &
- Earth BioGenome Project: Sequencing life for the future of life.
- 696 Proceedings of the National Academy of Sciences, 115(17): 4325-4333.
- 697 https://doi.org/10.1073/pnas.172011511

- 698 Lewin, H. A.; Richards, S.; Lieberman, A. E.; Allende, M. L.; Archibald, J. M.; Bálint, M.; ... &
- 699 Zhang, G. (2022). The earth BioGenome project 2020: Starting the clock. *Proceedings of*
- 700
 the
 National
 Academy
 of
 Sciences,
 119(4):
 e2115635118.

 701
 https://doi.org/10.1073/pnas.211563511
 https://doi.org/10.1073/pnas.211563511
 https://doi.org/10.1073/pnas.211563511
- 702 Lynggaard, C.; Calvignac-Spencer, S.; Chapman, C. A.; Kalbitzer, U.; Leendertz, F. H.; Omeja,
- P. A.; ... & Gogarten, J. F. (2023). Vertebrate environmental DNA from leaf swabs.
 Current Biology, 33(16): R853-R854. https://doi.org/10.1016/j.cub.2023.06.031
- 705 [dataset] Mammal Diversity Database. (2021). Mammal Diversity Database, v1.7. Zenodo.
 706 Access on 13/08/2023. http://doi.org/10.5281/zenodo.4139818
- Meek, P. D.; Vernes, K. & Falzon, G. (2013). On the reliability of expert identification of smallmedium sized mammals from camera trap photos. *Wildlife Biology in Practice*, 9(2): 461–
- 709 19. http://10.2461/wbp.2013.9.4
- 710 Mena, J. L.; Yagui, H.; Tejeda, V.; Bonifaz, E.; Bellemain, E.; Valentini, A.; ... & Lyet, A.
- (2021). Environmental DNA metabarcoding as a useful tool for evaluating terrestrial
 mammal diversity in tropical forests. *Ecological Applications*, 31(5): e02335.
 https://doi.org/10.1002/eap.2335
- 714 Mendes-Oliveira, A. C.; Peres, C. A.; Maues, P. C. R. A.; Oliveira, G. L.; Mineiro, I. G. B.;
- 715 Maria, S. L. S.; Lima, R. C. S. (2017) Oil palm monoculture induces drastic erosion of an
- 716 Amazonian forest mammal fauna. *PLoS ONE*, 12: e0187650.
 717 https://doi.org/10.1371/journal.pone.0187650
- 718 Monchamp, M. E.; Taranu, Z. E.; Garner, R. E.; Rehill, T.; Morissette, O.; Iversen, L. L.; ... &
- 719 Gregory-Eaves, I. (2023). Prioritizing taxa for genetic reference database development to

advance inland water conservation. *Biological Conservation*, 280: 109963.
https://doi.org/10.1016/j.biocon.2023.109963

- Moratelli, R. & Dias, D. (2015). A new species of nectar-feeding bat, genus *Lonchophylla*, from
 the Caatinga of Brazil (Chiroptera, Phyllostomidae). *ZooKeys*, (514): 73-91.
 https://doi.org/10.3897/zookeys.514.10013
- 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.
 https://doi.org/10.1038/35002501
- 728 Nakamura, A.; Kitching, R. L.; Cao, M.; Creedy, T. J.; Fayle, T. M.; Freiberg, M.; ... & Ashton,
- L. A. (2017). Forests and their canopies: achievements and horizons in canopy science.
 Trends in Ecology & Evolution, 32(6): 438-451. https://doi.org/10.1016/j.tree.2017.02.020
- 731 Paglia, A. P.; Da Fonseca, G. A.; Rylands, A. B.; Herrmann, G.; Aguiar, L. M.; Chiarello, A. G.;
- 732 ... & Patton, J. L. (2012). Lista Anotada dos Mamíferos do Brasil 2ª Edição/Annotated
 733 Checklist of Brazilian Mammals. *Occasional Papers in Conservation Biology*, 6: 1-82.
- Paknia, O.; Rajaei, S. H. & Koch, A. (2015). Lack of well-maintained natural history collections
 and taxonomists in megadiverse developing countries hampers global biodiversity
 exploration. *Organisms Diversity & Evolution*, 15: 619–629.
 https://doi.org/10.1007/s13127-015-0202-1
- 738 Pinto, I. D. S.; Chagas, B. D. D.; Rodrigues, A. A. F.; Ferreira, A. L.; Rezende, H. R.; Bruno, R.
- 739 V.; ... & Peixoto, A. A. (2015). DNA barcoding of neotropical sand flies (Diptera,
- 740 Psychodidae, Phlebotominae): species identification and discovery within Brazil. *PLoS*
- 741 ONE, 10(10): e0140636. https://doi.org/10.1371/journal.pone.0140636

- Pinto, F. A. S.; Cirino, D. W.; Cerqueira, R. C.; Rosa, C. & Freitas, S. R. (2022). How many
 mammals are killed on brazilian roads? Assessing impacts and conservation implications. *Diversity*, 14(10): 1-13. https://doi.org/10.3390/d14100835
- 745 Praill, L. C.; Eppley, T. M.; Shanee, S.; Cunneyworth, P. M. K.; Abra, F. D.; Allgas, N.; ... &
- 746 Svensson, M. S. (2023). Road Infrastructure and Primate Conservation: Introducing the
- Global Primate Roadkill Database. *Animals*, 13: 1-16. https://doi.org/10.3390/ani13101692
- 748 Quintela, F.; Da Rosa, C. A. & Feijo, A. (2020). Updated and annotated checklist of recent
- mammals from Brazil. Anais da Academia Brasileira de Ciências, 92: 1-57.
 https://doi.org/10.1590/0001-3765202020191004
- 751 Rengifo, E. M.; D'Elía, G.; García, G.; Charpentier, E. & Cornejo, F. M. (2022). A new species
- of titi monkey, genus Cheracebus Byrne et al., 2016 (Primates: Pitheciidae), from Peruvian
 Amazonia. *Mammal Study*, 48(1): 1-16. https://doi.org/10.3106/ms2022-0019
- Ripple, W. J.; Abernethy, K.; Betts, M. G.; Chapron, G.; Dirzo, R.; Galetti, M.; ... & Young, H.
- 755 (2016). Bushmeat hunting and extinction risk to the world's mammals. *Royal Society Open*756 *Science*, 3(10): 1-16. https://doi.org/10.1098/rsos.160498
- 757 Rosa, C. A.; Zenni, R.; Ziller, S. R.; de Almeida Curi, N. & Passamani, M. (2018). Assessing the
- risk of invasion of species in the pet trade in Brazil. *Perspectives in Ecology and Conservation*, 16(1): 38-42. https://doi.org/10.1016/j.pecon.2017.09.005
- Raven, P. H.; Gereau, R. E.; Phillipson, P. B.; Chatelain, C.; Jenkins, C. N. & Ulloa Ulloa, C.
- 761 (2020). The distribution of biodiversity richness in the tropics. *Science Advances*, 6(37):
 762 eabc6228. http://10.1126/sciadv.abc622
- 763 Rocha, R. G.; Duda, R.; Flores, T.; Rossi, R.; Sampaio, I.; Mendes-Oliveira, A. C.; ... & Costa,
- L. P. (2018). Cryptic diversity in the Oecomys roberti complex: revalidation of Oecomys

- tapajinus (Rodentia, Cricetidae). Journal of Mammalogy, 99(1): 174-186.
 https://doi.org/10.1093/jmammal/gyx149
- 767 Sanches, A.; Perez, W. A. M.; Figueiredo, M. G.; Rossini, B. C.; Cervini, M.; Galetti Jr, P. M. &

Galetti, M. (2011). Wildlife forensic DNA and lowland tapir (*Tapirus terrestris*) poaching.

- 769 Conservation Genetics Resources, 3(1): 189-193. https://doi.org/10.1007/s12686-010770 9318-y
- Sanches, A.; Tokumoto, P. M.; Peres, W. A.; Nunes, F. L.; Gotardi, M. S.; Carvalho, C. S.; ... &
 Galetti, M. (2012). Illegal hunting cases detected with molecular forensics in Brazil. *Investigative Genetics*, 3: 1-5. https://doi.org/10.1186/2041-2223-3-17
- 774 Saranholi, B. H.; Rodriguez-Castro, K. G.; Carvalho, C. S.; Chahad-Ehlers, S.; Gestich, C. C.;
- Andrade, S. C.; ... & Galetti Jr, 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
- 778 Scalon, M. C.; Domingos, F. M. C. B.; Cruz, W. J. A.; Jr, B. H. M.; Marimon, B. S. & Oliveras,
- I. (2020). Diversity of functional trade-offs enhances survival after fire in Neotropical
 savanna species. *Journal of Vegetation Science*, 31(1): 139–150.
 https://doi.org/10.1111/jvs.12823
- Schindel, D. E., & Miller, S. E. (2005). DNA barcoding a useful tool for taxonomists. *Nature*,
 435(17): 17. https://doi.org/10.1038/435017b
- 784 Scheffers, B. R.; Oliveira, B. F.; Lamb, I. & Edwards, D. P. (2019). Global wildlife trade across
- the tree of life. *Science*, 366(6461): 71-76. http://10.1126/science.aav532

- 786 Smart, U.; Cihlar, J. C. & Budowle, B. (2021). International Wildlife Trafficking: a perspective
- 787 on the challenges and potential forensic genetics solutions. *Forensic Science International:*788 *Genetics*, 54: 102551. https://doi.org/10.1016/j.fsigen.2021.102551
- Šmíd, J. (2022). Geographic and taxonomic biases in the vertebrate tree of life. *Journal of Biogeography*, 49(12): 2120-2129. https://doi.org/10.1111/jbi.14491
- 791 Theissinger, K.; Fernandes, C.; Formenti, G.; Bista, I.; Berg, P. R.; Bleidorn, C.; ... & Zammit,
- G. (2023). How genomics can help biodiversity conservation. *Trends in Genetics*, 39(7):
 545-559. https://doi.org/10.1016/j.tig.2023.01.005
- Therneau, T.; Atkinson, B. & Ripley, B. (2015). rpart: Recursive partitioning and regression
 trees. *R package version*, 4: 1-9.
- Threlfall, J. & Blaxter, M. (2021). Launching the tree of life gateway. *Wellcome Open Research*,
 6: 1-3. http://10.12688/wellcomeopenres.16913.1
- 798 Vergara-Asenjo, G.; Alfaro, F. M. & Pizarro-Araya, J. (2023). Linnean and Wallacean shortfalls
- in the knowledge of arthropod species in Chile: Challenges and implications for regional
- 800 conservation. *Biological Conservation*, 281: 1-10.
- 801 https://doi.org/10.1016/j.biocon.2023.110027
- 802 von Cräutlein, M.; Korpelainen, H.; Pietiläinen, M. & Rikkinen, J. (2011). DNA barcoding: a
- tool for improved taxon identification and detection of species diversity. *Biodiversity and*
- 804 *Conservation*, 20(2): 373-389. https://doi.org/10.1007/s10531-010-9964-0



Figure 1. Occurrences of DNA samples from Brazilian mammal species that are represented on
Bold Systems and the representativity of the mammal species that have been sequenced per
biome. AM= Amazon Forest, AF= Atlantic Forest, CE= Cerrado, PT= Pantanal, PP= Pampas,
MAR= Marine.



Figure 2. Distribution of mammal species richness in Brazil (left), distribution of Brazilian
mammal species richness with available sequences of DNA barcoding (center), and proportion of
Brazilian mammal species with available sequences of DNA barcoding in relation to the total
species richness of Brazilian mammals within an area (right).



Figure 3. Proportion of Brazilian mammals species represented on molecular databases
according to: A) species richness within each mammal order, B) conservation status, C) diet and
D, behavior.



Figure 4. Proportion of Brazilian mammals species represented on molecular databases
according to their taxonomic order and type of threat (poaching, roadkill, and illegal widlife
trade).



Figure S1. Distribution of mammal species richness of each order in Brazil (left panels), distribution of Brazilian mammal species richness with available sequences of DNA barcoding for each order (panels in the center), and proportion of Brazilian mammal species with available sequences of DNA barcoding in relation to the total species richnness of Brazilian mammals within an area for each order (right panels).



Figure S2. Results of Classification and Regression Trees analysis (CART analysis) to evaluate
the most important traits of Brazilian mammals that predict their probability of being sequenced
for the COI DNA barcode and deposited in BOLD systems.



Figure S3. Results of Classification and Regression Trees analysis (CART analysis) to evaluate
the most important traits of Brazilian mammals that predict their probability of being sequenced
for the COI DNA barcode and deposited in BOLD systems.

Priority	Order	Species	Ende-	Restric-	Conserva-	Population	Thre
			mic	ted distribu- tion	tion status	trend	at
EHP	Primates	Mico marcai	Yes	Yes	Vulnerable	Decreasing	IWT
EHP	Primates	Mico rondoni	Yes	Yes	Vulnerable	Decreasing	IWT
ЕНР	Primates	Alouatta ululata	Yes	Yes	Endangered	Decreasing	Р
EHP	Primates	Callicebus coimbrai	Yes	Yes	Endangered	Decreasing	IWT
EHP	Primates	Leontopithe- cus caissara	Yes	Yes	Endangered	Decreasing	P, IWT
EHP	Primates	Callithrix flaviceps	Yes	Yes	Critically endangered	Decreasing	P, IWT
EHP	Primates	Saguinus	Yes	Yes	Critically	Decreasing	P, R,

endangered

bicolor

857 Table SI. Prioritization for threatened Brazilian mammal species to be sequenced for DNA858 barcoding.

IWT

HP	Primates	Alouatta	Yes	No	Vulnerable	Decreasing	P, R,
		belzebu					IWT
HP	Primates	Alouatta	Yes	No	Vulnerable	Decreasing	None
		discolor					
HP	Primates	Alouatta	No	No	Vulnerable	Decreasing	P, R,
		guariba					IWT
HP	Primates	Alouatta	No	No	Vulnerable	Decreasing	Р
		puruensis					
HP	Primates	Callicebus	Yes	No	Vulnerable	Decreasing	P,
		melanochir					IWT
HP	Primates	Callicebus	No	No	Vulnerable	Decreasing	P,
		brunneus					IWT
HP	Primates	Cebuella	No	No	Vulnerable	Decreasing	P,
		pygmaea					WT
HP	Primates	Chiropotes	Yes	No	Vulnerable	Decreasing	P,
		utahickae					IWT
HP	Rodentia	Euryoryzo-	Yes	Yes	Vulnerable	Decreasing	None
		mys lamia					
HP	Primates	Lagothrix	No	No	Vulnerable	Decreasing	Р
		lagothrixa					

HP	Didelphi-	Marmosa	No	No	Vulnerable	Decreasing	None
	morphia	phaea					
HP	Cetartio-	Mazama	Yes	No	Vulnerable	Unknown	P, R
	dactyla	nana					
HP	Primates	Mico	Yes	No	Vulnerable	Decreasing	None
		munduruku					
HP	Primates	Saguinus	Yes	No	Vulnerable	Decreasing	None
		ursula					
HP	Primates	Sapajus cay	No	No	Vulnerable	Decreasing	Р
HP	Lagomor-	Sylvilagus	Yes	Yes	Vulnerable	Unknown	None
	pha	tapetillus					
HP	Sirenia	Trichechus	No	No	Vulnerable	Decreasing	Р,
		inunguis					IWT
HP	Primates	Ateles	Yes	No	Endangered	Decreasing	P, R,
		marginatus					IWT
HP	Primates	Callithrix	Yes	No	Endangered	Decreasing	R,
		aurita					IWT
HP	Primates	Chiropotes	Yes	No	Endangered	Decreasing	P,
		satanas					IWT

HP	Rodentia	Coendou speratus	Yes	Yes	Endangered	Decreasing	None
НР	Primates	Lagothrix poeppigii	No	No	Endangered	Decreasing	P, IWT
HP	Chiropte- ra	Lonchophylla bokermanni	Yes	Yes	Endangered	Unknown	None
HP	Chiropte- ra	Lonchophylla dekeysery	Yes	Yes	Endangered	Decreasing	None
НР	Primates	Mico schneideri	Yes	No	Endangered	Decreasing	None
HP	Rodentia	Microako- dontomys transitorius	Yes	Yes	Endangered	Decreasing	None
НР	Rodentia	Phyllomys thomasi	Yes	Yes	Endangered	Decreasing	None
НР	Primates	Sapajus robustus	Yes	No	Endangered	Decreasing	R
HP	Primates	Brachyteles hypoxanthus	Yes	No	Critically endangered	Decreasing	P, IWT

HP	Primates	Callicebus	Yes	No	Critically	Decreasing	P,
		barbarabrow-			endangered		IWT
		nae					
HP	Primates	Callicebus	Yes	No	Critically	Decreasing	None
		grovesi			endangered		
HP	Primates	Callicebus	Yes	No	Critically	Decreasing	IWT
		vieirai			endangered		
HP	Rodentia	Cavia	Yes	Yes	Critically	Decreasing	None
		intermedia			endangered		
HP	Didelphi-	Monodelphis	Yes	Yes	Critically	Unknown	None
	morphia	unistriata			endangered		
HP	Rodentia	Phyllomys	Yes	Yes	Critically	Decreasing	None
		unicolor			endangered		
IP	Primates	Cebuella	No	No	Vulnerable	Decreasing	None
		niveiventris					
IP	Primates	Cebus	No	No	Vulnerable	Decreasing	None
		unicolor					
IP	Rodentia	Gyldenstol-	No	No	Critically	Unknown	None
		pia fronto			endangered		

^{859 *}P (poaching), R (roadkill), IWT (illegal wildlife trade).

860	**EHP (Extremely high priority), HP (high priority), IP (intermediate priority).
861	
862	
863	
864	
865	
866	
867	
868	
869	
870	
871	
872	
873	