Identification of non-model mammal species using the MinION DNA sequencer from Oxford Nanopore

Sara Velasquez-Restrepo[1], Mariana Corrales Orozco[1], Nicolás D. Franco-Sierra[2,3], Juan M. Martínez-Cerón[1] and Juan F. Díaz-Nieto[1]

[1] Natural Systems and Sustainability Area, Universidad EAFIT, Medellín, Antioquia, Colombia
[2] Syndesis Health, Palm Beach Gardens, Florida, United States
[3] Corporación de Investigación e Innovación (VEDAS CII), VEDAS, Medellín, Antioquia, Colombia

ABSTRACT

Background: The Neotropics harbors the largest species richness of the planet; however, even in well-studied groups, there are potentially hundreds of species that lack a formal description, and likewise, many already described taxa are difficult to identify using morphology. Specifically in small mammals, complex morphological diagnoses have been facilitated by the use of molecular data, particularly from mitochondrial sequences, to obtain accurate species identifications. Obtaining mitochondrial markers implies the use of PCR and specific primers, which are largely absent for non-model organisms. Oxford Nanopore Technologies (ONT) is a new alternative for sequencing the entire mitochondrial genome without the need for specific primers. Only a limited number of studies have employed exclusively ONT long-reads to assemble mitochondrial genomes, and few studies have yet evaluated the usefulness of such reads in multiple non-model organisms.

Methods: We implemented fieldwork to collect small mammals, including rodents, bats, and marsupials, in five localities in the northern extreme of the Cordillera Central of Colombia. DNA samples were sequenced using the MinION device and Flongle flow cells. Shotgun-sequenced data was used to reconstruct the mitochondrial genome of all the samples. In parallel, using a customized computational pipeline, species-level identifications were obtained based on sequencing raw reads (Whole Genome Sequencing). ONT-based identifications were corroborated using traditional morphological characters and phylogenetic analyses.

Results: A total of 24 individuals from 18 species were collected, morphologically identified, and deposited in the biological collection of Universidad EAFIT. Our different computational pipelines were able to reconstruct mitochondrial genomes from exclusively ONT reads. We obtained three new mitochondrial genomes and eight new molecular mitochondrial sequences for six species. Our species identification pipeline was able to obtain accurate species identifications for up to 75% of the individuals in as little as 5 s. Finally, our phylogenetic analyses corroborated the identifications from our automated species identification pipeline and revealed important contributions to the knowledge of the diversity of Neotropical small mammals.

Discussion: This study was able to evaluate different pipelines to reconstruct mitochondrial genomes from non-model organisms, using exclusively ONT reads,

Submitted 21 December 2023
Accepted 18 July 2024
Published 25 September 2024

Corresponding author
Juan F. Díaz-Nieto,
jdiazni@eafit.edu.co

Academic editor
Richard Schuster

Additional Information and Declarations can be found on page 27

DOI **10.7717/peerj.17887**

Copyright
2024 Velasquez-Restrepo et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

How to cite this article Velasquez-Restrepo S, Corrales Orozco M, Franco-Sierra ND, Martínez-Cerón JM, Díaz-Nieto JF. 2024. Identification of non-model mammal species using the MinION DNA sequencer from Oxford Nanopore. **PeerJ 12:e17887**
DOI 10.7717/peerj.17887

benchmarking these protocols on a multi-species dataset. The proposed methodology can be applied by non-expert taxonomists and has the potential to be implemented in real-time, without the need to euthanize the organisms and under field conditions. Therefore, it stands as a relevant tool to help increase the available data for non-model organisms, and the rate at which researchers can characterize life specially in highly biodiverse places as the Neotropics.

Subjects Biodiversity, Bioinformatics, Molecular Biology, Zoology
Keywords Species identification, Cryptic species, Non-model organisms, Mitochondrial genomes, Computational pipeline, Shotgun sequencing, Third generation sequencing

INTRODUCTION

The Neotropics holds some of the greatest species richness and endemism in the world; however, a large fraction of such diversity is still unknown and hidden within "cryptic diversity" (*Field et al., 2009*; *Antonelli & Sanmartín, 2011*; *Moreno et al., 2018*; *Parsons et al., 2022*). For the past three decades, the recognition and description of metazoan cryptic species has been based on the complementary use of DNA sequences (with emphasis on mitochondrial DNA) and extremely detailed morphological characters (*Byrne, 2023*; *Nachman et al., 2023*). Particularly for small mammals (rodents, bats, opossums), the diagnostic characters—traits unique to a species which are vital for its identification—are almost entirely based on cranial, dental, or soft-tissue morphology that inevitably requires the collection (euthanization) of animals to be prepared as study specimens (*Byrne, 2023*; *Nachman et al., 2023*). Nevertheless, euthanizing individuals may be highly undesirable, especially in conservation scenarios such as the presence of endangered species (*Kyle & Wilson, 2007*; *Cleman et al., 2014*; *Costello et al., 2016*).

DNA sequencing has emerged as a vital tool to overcome the challenges associated with species identification solely based on morphological traits (*Bickford et al., 2007*; *Schlegel et al., 2012*; *Lopez-Baucells et al., 2018*), and in particular, mitochondrial DNA (mtDNA) has become the ideal source of data for species identification, given to its—compared to the nuclear DNA—relatively short coalescent time, small effective population size, and high mutation rate (*Brown, George & Wilson, 1979*; *Moore, 1995*). Although, within this genome, the Cytochrome Oxidase C Subunit I (COI) has gained prominence for the identification of a wide range of Metazoan organisms (*Gostel & Kress, 2022*; *Hebert, Ratnasingham & de Waard, 2003*; *Hebert et al., 2003*), other various mitochondrial markers (*e.g.*, 12S, 16S, CYTB, NAD) have been historically used for species identification in particular clades (for some examples see *Parson et al., 2000*; *Rastogi et al., 2007*; *Milan et al., 2020*). Consequently, these multiple mtDNA markers have produced heterogeneous sequence databases (*Kowalczyk et al., 2021*; *Antil et al., 2022*), which pose a great challenge for species identification because of the absence of complete datasets of orthologues needed for comparison.

Obtaining particular amplicons for a molecular-based species identification is still commonly done using PCR-based amplification (*e.g.*, *Nowak et al., 2014*; *Schäffer, Zachos & Koblmüller, 2017*; *Pomerantz et al., 2018*, *2022*). However, although very useful, PCR

requires the use of specific primers which can be absent for multiple non-model organisms (*Schäffer, Zachos & Koblmüller, 2017*; *Cicero et al., 2021*). In many taxonomic groups, there are primer cocktails designed for a vast range of species, but the construction of such "universal primers" relies on genomes already available in databases (usually from model organisms), where there is limited representation of Neotropical genetic diversity (*Naidu et al., 2012*; *Schlegel et al., 2012*; *da Fonseca et al., 2016*). Hence, in the presence of understudied species, the use of these primers can lead to the inability to obtain the DNA fragment of interest, problems of specificity, or obtaining the wrong DNA fragment (*Naidu et al., 2012*; *Schlegel et al., 2012*; *da Fonseca et al., 2016*).

One way to deal with the heterogeneity of mtDNA databases is to sequence the entire mitochondrial genome and have all the potential markers for the species identification regardless of the marker available in the DNA repository. Sequencing the mitogenome has been traditionally achieved through long-range PCR (*Groenenberg et al., 2017*), back-to-back PCRs (*Emser et al., 2021*) and shotgun or targeted massive parallel sequencing (*Zhang, Cui & Wong, 2012*; *Clarke et al., 2014*). Although, any of these methods has been sufficient for obtaining (routinely) mitochondrial genomes, it requires sequencing platforms (*e.g.*, Sanger, Illumina) and molecular biology facilities which are often unavailable in developing countries (*Ciocca & Delgado, 2017*; *Galina, Martínez & Murphy, 2023*). Consequently, a typical sequencing approach involves exporting the DNA or amplicons to sequencing facilities, adding paperwork, time and budget to this process (*Salager-Meyer, 2008*; *Rios & de Regules, 2022*; *Zarog, 2023*). The cumulative impact of the mentioned problems hinders accurate species identification using DNA sequences, affecting the efficiency of data collection (*i.e.*, DNA sequences) and the expansion of databases (*Rios & de Regules, 2022*; *Zarog, 2023*).

The development of the third-generation sequencing platform of Oxford Nanopore Technologies (ONT), has resulted as a promising alternative to obtain sequences on-site without the need of complex facilities. In fact, the portable MinION sequencing platform has already been used to shotgun-sequence—without the need of primers—complete mitochondrial genomes in multiple taxonomic groups such as invertebrates (*Gan et al., 2019*; *Baeza, 2020*; *Filipović et al., 2022*), vertebrates (*Wiley & Miller, 2020*; *Kryukov et al., 2020*; *Malukiewicz et al., 2021*; *Baeza & García-De León, 2022*), fungi (*Sa et al., 2020*) and plants (*Jackman et al., 2020*). Furthermore, ONT has emerged as a viable alternative for species identification, primarily within a DNA barcoding framework, which involves the sequencing of PCR-based amplicons, and it has already proven to be effective in diverse taxonomic groups including bacteria and viruses (*Benítez-Páez, Portune & Sanz, 2016*; *Riaz et al., 2021*), plants (*Antil et al., 2022*), invertebrates (*Kneubehl et al., 2022*) and vertebrates (*Krehenwinkel, Pomerantz & Prost, 2019*; *Srivathsan et al., 2019*, *2021*; *Johri et al., 2019*). Notably, in recent years a few studies have shown the capabilities of this sequencing technology for real-time and onsite species identification (*Parker et al., 2017*; *Pomerantz et al., 2018*, *2022*; *Maestri et al., 2019*).

Some studies have already used ONT shotgun sequencing reads for species identification, offering an alternative approach to circumvent the challenges associated

with the amplification of specific markers (*Baeza, 2020*; *Imai et al., 2020*; *Baeza & García-De León, 2022*; *De Vivo et al., 2022*). These studies mostly provide identifications implementing phylogenetic analyses, but at least one study uses ONT reads followed by a custom-made bioinformatic pipeline to provide species-level identification of a non-model mammal species (*Franco-Sierra & Díaz-Nieto, 2020*). Despite the advancements that these projects made in the field of ONT-based species identification, they usually include one or few individuals of closely related species, and they have failed to test their laboratory and computational methods with multiple individuals across a broader taxonomical sampling (different families or orders). Also, these studies rely on expensive flowcells (ca. 900 USD) with massive sequencing capacity, which although generate the amount of data needed for species identification, the benefit to cost ratio for a small mitochondrial genome (ca. 17 kb) can be considered extravagant. Currently, Oxford Nanopore Technologies have produced a small-essay sequencing flowcell, the Flongle, which is available at a reduced cost (compared to the regular flow cells) and offers a remarkable level of precision with a mere 5% chemical error rate in the R9 series. Our research aims to evaluate the feasibility of using ONT shotgun reads obtained with the Flongle sequencing flowcell for the identification of multiple species of non-model organism of mammals.

MATERIALS AND METHODS

Fieldwork

We conducted fieldwork in five different localities near the city of Medellín so we could rapidly process tissue samples of captured specimens. These localities include cloud mountain forest, low tropical mountain forest, tropical dry forest, and urban ecosystems (Fig. 1 and Appendix 1). We sampled each locality for 8 days implementing the trapping methodologies specific for each focal group (bats, rodents, and marsupials) proposed by *Voss & Emmons (1996)*. For volant mammals (bats) we used eight nets of three different sizes (6 m × 3 m, 9 m × 3 m, and 12 m × 3 m) per night. For terrestrial small mammals, at each locality, we used 100 standard Sherman folding live traps (80 cm × 90 cm × 230 cm) per night. Finally, we baited traps every morning with a mixture of oats, banana, and vanilla essence, or with a single slice of banana.

To validate the results obtained from our species identification pipeline (see section "Species identification pipeline with ONT raw reads"), it is essential to compare them against an accurate species identification, which in our case is based on morphology and phylogenetic systematics. Therefore, it was still necessary to collect the individuals and deposit them as voucher specimens in a museum collection. We followed all the methods of capture and handling approved by the American Society of Mammalogists Animal Care and Use Committee (*Sikes & Bryan, 2016*). Specimens selected for collection were sacrificed using cervical dislocation or exposure to high concentrations of CO_2 (carbon dioxide displacement rate of 60% volume per minute over a duration of 2 min), in accordance with the AVMA Guidelines for the euthanasia of animals: 2020 edition (*American Veterinary Medical Association (AVMA), 2020*). Cervical dislocation was used primarily for bats, while terrestrial mammals were euthanized with CO_2. For the

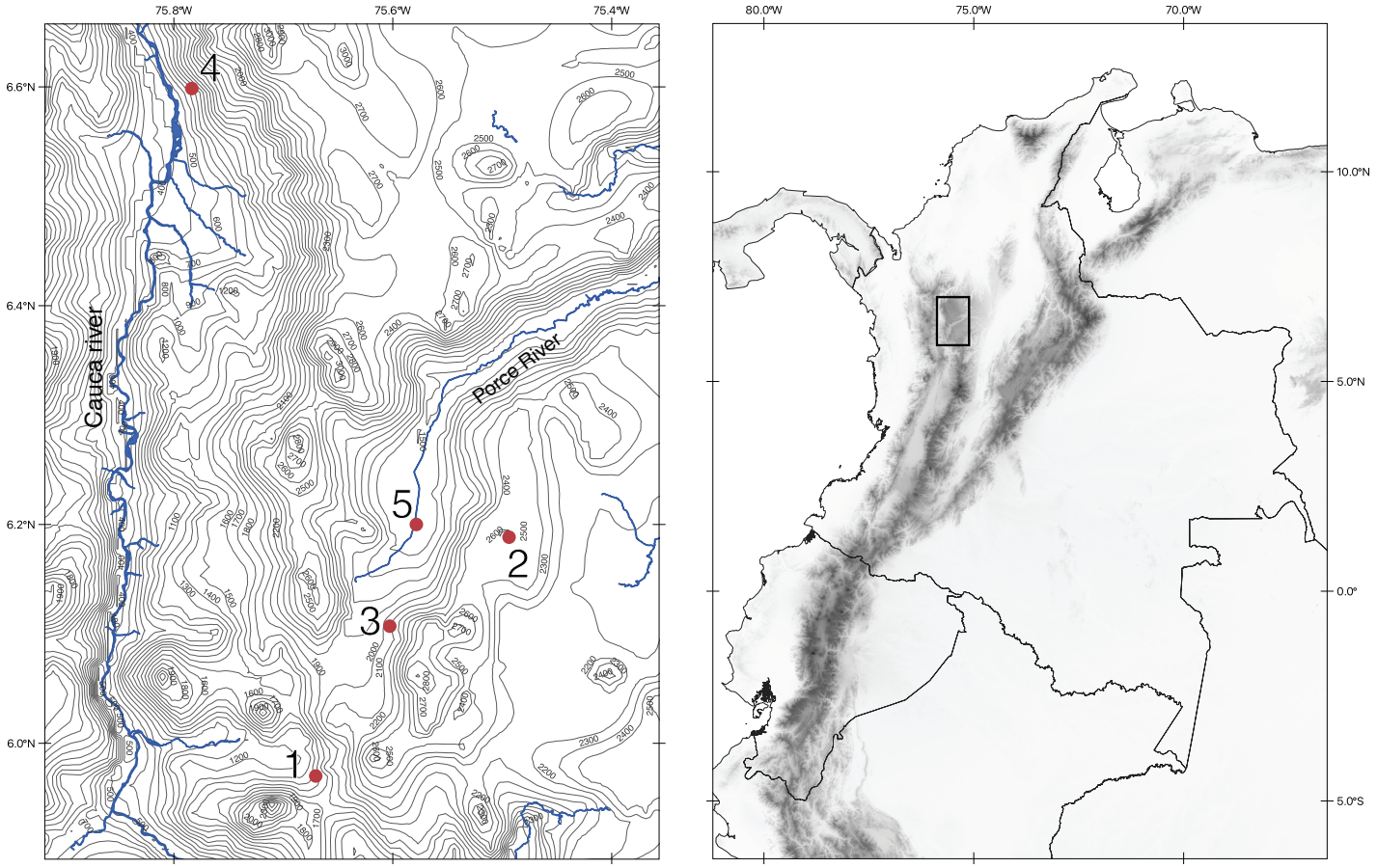

Figure 1 Map of the northern section of the Cordillera Central in Colombia showing the localities of our fieldwork. Map made by the authors, using the QGIS program (https://www.qgis.org/en/site/index.html) and spatial data from DIVA-GIS (http://diva-gis.org/). Both repositories are Free and Open Source Geographic Information System. Full-size DOI: 10.7717/peerj.17887/fig-1

morphological identification, we used specialized literature for each group as needed. All the material collected is deposited at the biological collection of Universidad EAFIT (Medellín, Colombia). All the methodological procedures of this project were approved by the Ethics Committee of the Universidad EAFIT. The approval of the field work permit was granted by the National Environmental Licensing Authority (ANLA) through Resolution No. 1160 of 2023.

Laboratory work and Oxford Nanopore Sequencing

For the DNA extraction we removed six millimeters from the distal section of the tail of rodents and opossums, and three patagium (wing membrane) samples with a 3 mm dermatological punch for bats. We washed each tissue sample with molecular biology grade ethanol to remove external contaminants. Subsequently, we performed whole genome DNA extraction using the DNeasy Blood and Tissue Kit (Qiagen Venlo, The Netherlands) following manufacturer's instructions, and measured DNA concentration by fluorometry using the dsDNA BR assay from Qubit 3.0 (Thermo Fisher Scientific,

Waltham, MA, USA). Next, for Whole Genome Sequencing, we used the MinION device with Flongle flow cells (FLO-FLG001, R9.4.1 chemistry). Library preparation was the same for all samples using the Ligation Sequencing Kit SQK-LSK110 following the Genomic DNA by Ligation SQK-LSK110-Flongle Nanopore Protocol (version GDE_9108_v110_revL_10Nov2020). Noteworthy, we conducted the adapter ligation clean-up step using LFB buffer to enhance the recovery of longer DNA fragments (>3 Kb). For the library loading step, we used 6 µl of DNA library in addition to the sequencing Buffer II (SBII) and Loading Beads II (LBII) from the Flongle Expansion kit which are stored in glass vials. We loaded the library (sequencing mix) dropwise onto the sample port ensuring that the mix entered the Flongle cell by capillary action. We controlled the MinION sequencing run from a Dell laptop (processor Intel® Xeon® CPU E3-1505M v5 @ 2.8 Hz × 4 and 32 Gb ram) using the MinKNOW software v21.05 setting a runtime of 24 h for each sample without live basecalling. After sequencing, we basecalled the raw signal (FAST5 files) using the ONT Guppy Sequencing Pipeline software 5.0.11 to obtain sequencing reads in FASTQ format. Subsequently, we used the sequencing summary file from the basecalling process to obtain sequencing performance statistics using NanoPlot 1.39 (*De Coster et al., 2018*). We filtered the data from each run by quality with NanoFilt 2.8.0 (*De Coster et al., 2018*) and only retained the reads with Q-score ≥ 7.

Mitogenome reconstruction

The reconstruction of mitochondrial genomes was conducted under two distinct scenarios (Fig. 2). First, when there is no prior knowledge about the species identity of the organism, we implemented the mitochondrial genome reconstruction based on Whole Genome Sequences. Second, when there is prior information available on the identification of the organism (*e.g.*, order or family)—and more importantly—when there is availability of closely related reference mitogenomes in databases, we used a mitochondrial-filtered approach. For this latter approach we filtered the mitochondrial reads implementing both Minimap2 (*Li, 2018*) and the mtBlaster script (*Franco-Sierra & Díaz-Nieto, 2020*). For both scenarios, we applied two distinct assembly methods: a *de novo* assembler implemented in Flye 2.9 (*Kolmogorov et al., 2019*), and a reference-based assembler implemented in Rebaler 0.2.0 (available at https://github.com/rrwick/Rebaler). We used the –meta flag for Flye which has been designed for metagenome/uneven coverage assemblies and is able to recover different genomes within the same sample, an ideal fit for the mitochondrial/nuclear genome uneven scenario obtained with the WGS data (*Kolmogorov et al., 2019*, *2020*). As Rebaler is a reference-based assembler, we used as reference the mitochondrial genome from the most closely related species available in the NCBI. This selection was based on the initial morphological identifications of each species. In total, we implemented six (6) different routines based on the data type (WGS and mtDNA-filtered) and assembly method (Flye and Rebaler) (Fig. 2). The first two pipelines started from WGS data, referred to as WGS (Fig. 2, pipelines 1 and 2), and a second set of pipelines used mitochondrial-filtered reads, referred to as Minimap2-filtered and mtBlaster-filtered (Fig. 2, pipelines 3–6). The name given to the resulting mitochondrial genome assemblies indicates the data type and assembler used to reconstruct them. For

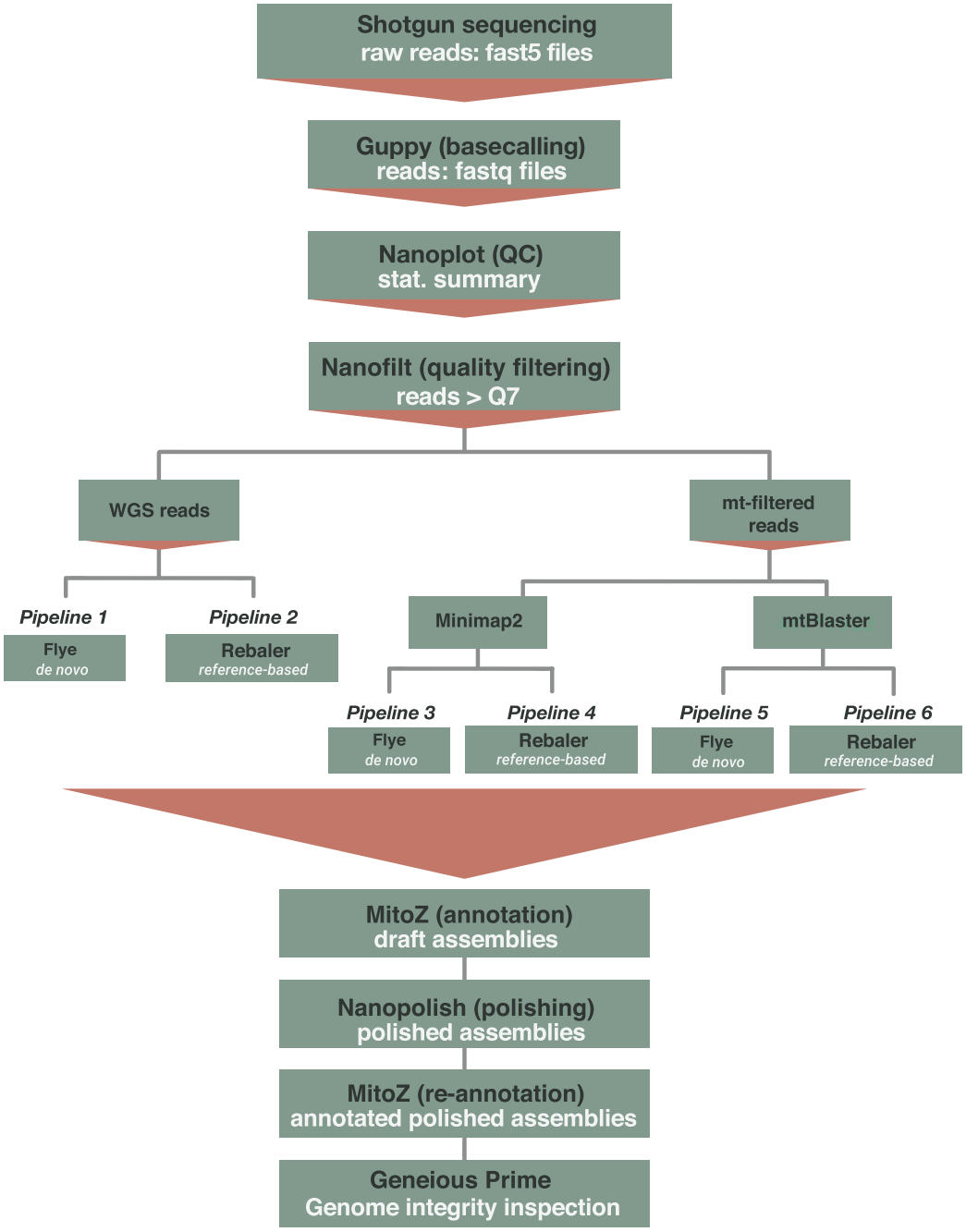

Figure 2 Computational workflow implemented for the reconstruction of mitochondrial genomes from ONT shotgun sequencing data. Pipelines 1 and 2 use WGS reads and use *de novo* (Flye) and reference based (Rebaler) assemblers. Pipelines 3 to 6, initially filter the mitochondrial reads implementing Minimap2 (pipelines 3 and 4) or mtBlastr (pipeline 5 and 6) to subsequently execute the assemble step (Flye and Rebaler). The final steps (below red triangle) that include annotation and polishing are common to all six pipelines. Full-size DOI: 10.7717/peerj.17887/fig-2

example, if a genome was reconstructed using mitochondrial data filtered with Minimap2 and subsequently assembled with Flye, the resulting genome would be labeled as Minimap2+Flye.

We annotated the resulting assemblies (final assembly generated by either Flye or Rebaler, "draft assembly") using the command line version of MitoZ v2.4.1 (*Meng et al., 2019*), then we polished the annotated genomes using NanoPolish 0.14.0 (*Loman, Quick & Simpson, 2015*), and finally re-annotated the polished genomes ("annotated polished assemblies" in Fig. 2). We evaluated the quality of all the annotated mitogenomes (polished and not polished) using two distinct assembly criteria: hits and errors. Assembly hits correspond to complete coding genes with open reading frames (complete genes) and complete non-coding genes (complete rRNA, tRNA). On the other hand, assembly errors are coding genes with premature stop codons (denoted as stop codons), incomplete genes (incomplete), absent genes within an assembly (null), and not assembled genomes (not assembled). We aligned each mitochondrial genome assembly to a closely related reference genome (the same used for Rebaler) to evaluate the hit and error criteria (Tables S2–S5).

Species identification based on phylogenetic analyses

Because it is essential to demonstrate the accuracy of the identifications obtained with the species identification pipeline (see below), it was necessary to provide a reliable identification with an independent line of evidence, which in our case was done with phylogenetic inference analyses. For this purpose, we extracted the mitochondrial DNA markers COI and Cytb—which are widely implemented for species identification in mammals (*Parson et al., 2000*; *Hebert, Ratnasingham & de Waard, 2003*)—from each of the annotated mitochondrial genomes, but we only extracted those sequences with sufficient quality (*i.e.*, Open Reading Frame, ORF). Then, we downloaded all the available sequences from GenBank for each of these two loci and constructed datasets for each genus, consequently, each dataset includes as ingroup all the GenBank sequences available for a specific marker within a particular genus, along with our extracted sequences (Table S1, File S2). We removed all duplicate sequences with the Dedupe algorithm implemented in Geneious 9.1.8 (*Kearse et al., 2012*) and used a sequence from a species of the sister genus as outgroup (Table S2). We aligned each dataset using MUSCLE (*Edgar, 2004*) and conducted independent gene tree analyses for each locus (COI and Cytb) based on a Maximum Likelihood approach. The best-fitting nucleotide substitution model for each alignment was determined using the Bayesian Information Criterion (BIC) and nodal support was evaluated with 1,000 bootstrap replicates using the Ultrafast Bootstrap Approximation (*Hoang et al., 2018*). All the phylogenetic analyses were implemented in the IQ-TREE software v2.2.0 beta COVID-edition (*Nguyen et al., 2015*). The resulting trees were visualized and edited using iTOL V5 (*Letunic & Bork, 2021*).

Species identification pipeline with ONT raw reads

For an automated species identification using the ONT raw reads, we followed the methodological approach of *Franco-Sierra & Díaz-Nieto (2020)* with some minor modifications. First, to obtain exclusively mtDNA reads from our WGS, we ran a blastn search using BLAST 2.13 (*Camacho et al., 2009*) implemented in the mtBlaster pipeline (https://github.com/nidafra92/squirrel-project/blob/master/mtblaster.py) using a

mitochondrial genome database of ten model organisms of Metazoa (see *Franco-Sierra & Díaz-Nieto, 2020*). We defined an identity cutoff value of 90% based on an identity threshold of 95% to safeguard homology at the nucleotide level plus the average ONT error rate of ~5% (for the R9 ONT chemistry). Subsequently, we blasted the extracted mitochondrial reads against a new custom-made database containing all (6,620,011) metazoan mitochondrial sequences available in GenBank as of March 2023. For the BLAST search, the strictness of the analysis ranged between medium to high level by setting the parameters as follows: max_target_seqs =1, word_size = 11, gapopen = 1, gapextend = 2, penalty = −4, reward = 1, max_hsps = 1, perc_identity = 90, and task = blastn.

We simulated three different hypothetical scenarios, where the size of the search database (custom database of mitochondrial sequences) varied depending on the amount of taxonomic information available for each individual: (a) when no information is available (complete Metazoa mtDNA dataset), (b) when sample identity is known at the order level, and (c) when sample identity is known at the family level. For each individual we conducted three independent searches, simulating all three scenarios (Metazoa, order, and family). To enhance the time-efficiency of BLAST searches, we implemented the program DC-BLAST (*Yim & Cushman, 2017*). For the order and family-level search scenarios, we refined the complete Metazoa database informing the DC-BLAST program with a text file containing a comprehensive list of NCBI Taxonomy IDs, specific to each particular taxonomic level. Briefly, for flying mammals all specimens are part of the order Chiroptera and families Molossidae, Phyllostomidae, and Vespertilionidae (Table 1); non-flying small mammals are contained in the orders Didelphimorphia and Rodentia, and families Didelphidae and Cricetidae, respectively (Table 1). For each level of analysis, we also evaluated the time efficiency of DC-BLAST. We implemented all species identification analyses in the computational cluster Apolo Scientific Computing Center (Universidad EAFIT) using the DC-BLAST (Divide and Conquer BLAST for HPC) utility (*Yim & Cushman, 2017*). The setup used for executions consisted of two HPC nodes with 32 cores per node, 64 GB RAM per node, Intel® Xeon® CPU E5-2683 v4 @ 2.10 GHz processor. After obtaining the DC-BLAST results for all experiments, we ranked the hits for each individual (reads from each mammal collected) at each level of analysis (Metazoa, order, family) using two key parameters: first in ascending order of their e-values, followed by percentage of identity in descending order (File S1). We consider the first hit in each analysis as the most reliable identification.

RESULTS

Field work, laboratory work and Oxford Nanopore Sequencing

We collected a total of 24 individuals, representing 18 species, five families, and three orders (Table 1). Our DNA extractions produced an average concentration of 35.7 ng/µl (7–102 ng/µl; 17 samples) for bat patagium tissue, and 69.85 ng/µl (26.6–1,147 ng/µl; seven samples) for terrestrial mammals. We found that library preparation reduced the DNA concentration by approximately 53.81% for most samples; however, in some instances this

Table 1 Morphological identification, locality, DNA extraction characteristics, and ONT sequence performance statistics of all specimens sequenced in this study.

EAFIT collection number	Morphological ID	Locality	Tissue	Concentration of DNA extraction (ng/µl)	Concentration of DNA library (ng/µl)	Number of cell pores	Sequencing run yield (Mbp)	Mean read length (Kbp)	Number of reads	Number of reads ≥ Q–7	% Reads ≥ Q–7	% mtDNA (Minimap2)	% mtDNA (mtBlaster)
EAFIT-M 2126	*Anoura caudata*	Hacienda la Sierra	P (2)	13	7.44	1	6.96	7.24	962	744	77.34	–	–
EAFIT-M 2106	*Myotis riparius*	Hacienda la Sierra	P (2)	7	21.4	54	265.13	5.23	50,689	33,947	66.97	0.02	0.02
EAFIT-M 2143	*Nephelomys pectoralis*	Hacienda la Sierra	T	33.8	24.2	38	251.67	3.92	64,068	56,059	87.5	0.07	0.08
EAFIT-M 2107	*Histiotus humboldtii*	Hacienda la Sierra	P (2)	11.4	6.88	85	133.35	3.73	35,714	30,019	84.05	0.06	0.04
EAFIT-M 2124	*Rhipidomys latimanus*	Hacienda la Sierra	T	76	21.6	66	720.86	2.96	243,742	212,128	87.03	0.05	0
EAFIT-M 2116	*Handleyomys alfaroi*	Hacienda la Sierra	T	78.2	17.3	45	195.09	2.6	75,000	57,366	76.49	0.03	0.04
EAFIT-M 2160	*Sturnira bogotensis*	La Miel	P (2)	13.7	7.34	55	615.41	3.04	202,206	170,523	84.33	0.02	0.04
EAFIT-M 2157	*Carollia brevicauda*	La Miel	P (3)	39.8	6.66	57	277.84	2.86	88,680	69,111	77.93	0	0.03
EAFIT-M 2164	*Anoura geoffroyi*	La Miel	P (3)	54.4	51	69	752.48	2.97	253,665	208,819	82.32	0.09	0.08
EAFIT-M 2158	*Artibeus lituratus*	Universidad EAFIT	P (3)	102	9.86	70	271.02	2.3	118,000	101,131	85.7	0.04	0.03
EAFIT-M 2167	*Glossophaga soricina*	Universidad EAFIT	P (3)	22	42	18	178.15	3.64	51,355	38,791	75.54	0	0
EAFIT-M 2166	*Artibeus lituratus*	Universidad EAFIT	P (3)	43.4	63.6	–	653.56	2.66	245,545	185,909	75.71	0.02	0.02
EAFIT-M 2165	*Sturnira bogotensis*	La Miel	P (3)	58.6	34.8	39	258.84	3.15	82,287	76,768	93.29	0	0.03
EAFIT-M 2168	*Carollia brevicauda*	La Miel	P (3)	28.6	33	88	562.47	2.95	190,895	152,744	80.01	0.01	0.02
EAFIT-M 2151	*Nephelomys childi*	Finca Cyathea	T	114	29.6	94	921.72	2.08	443,460	370,536	83.56	0.08	0.06
EAFIT-M 2159	*Myotis caucensis*	Universidad EAFIT	P (3)	16.2	15.4	76	535.48	3.8	140,848	116,425	82.66	0.01	0.01
EAFIT-M 2150	*Nephelomys childi*	Finca Cyathea	T	41	49.8	92	920.81	3.97	231,989	194,087	83.66	0.09	0.06
EAFIT-M 2152	*Molossus molossus*	Universidad EAFIT	P (3)	41	69.6	92	701.68	3.81	184,279	152,350	82.67	0.1	0.06
EAFIT-M 2161	*Akodon affinis*	Finca Cyathea	T	26.6	42.8	90	755.79	1.66	456,585	357,353	78.27	0.04	0.06
EAFIT-M 2153	*Artibeus lituratus*	Finca Socolao	P (3)	37.5	7.06	0	0	0	0	–	–	–	–
EAFIT-M 2154	*Phyllostomus discolor*	Finca Socolao	P (3)	46.9	8.2	0	0	0	0	–	–	–	–
EAFIT-M 2156	*Marmosa isthmica*	Finca Socolao	T	70.8	27.2	26	150.51	2.53	59,398	48,630	81.87	0.38	0.28
EAFIT-M 2155	*Marmosa isthmica*	Finca Socolao	T	84.2	65.4	72	523.36	3.33	157,390	123,681	78.58	0.41	0.28
EAFIT-M 2162	*Zygodontomys brevicauda*	Finca Socolao	T	104	40.4	55	530.42	3.03	175,153	142,932	81.6	0.04	0.03

Note:
Under the column Tissue samples, P makes reference to bat patagium and the number in parentheses are the dermatological punches sampled; T corresponds to tail of terrestrial mammals.

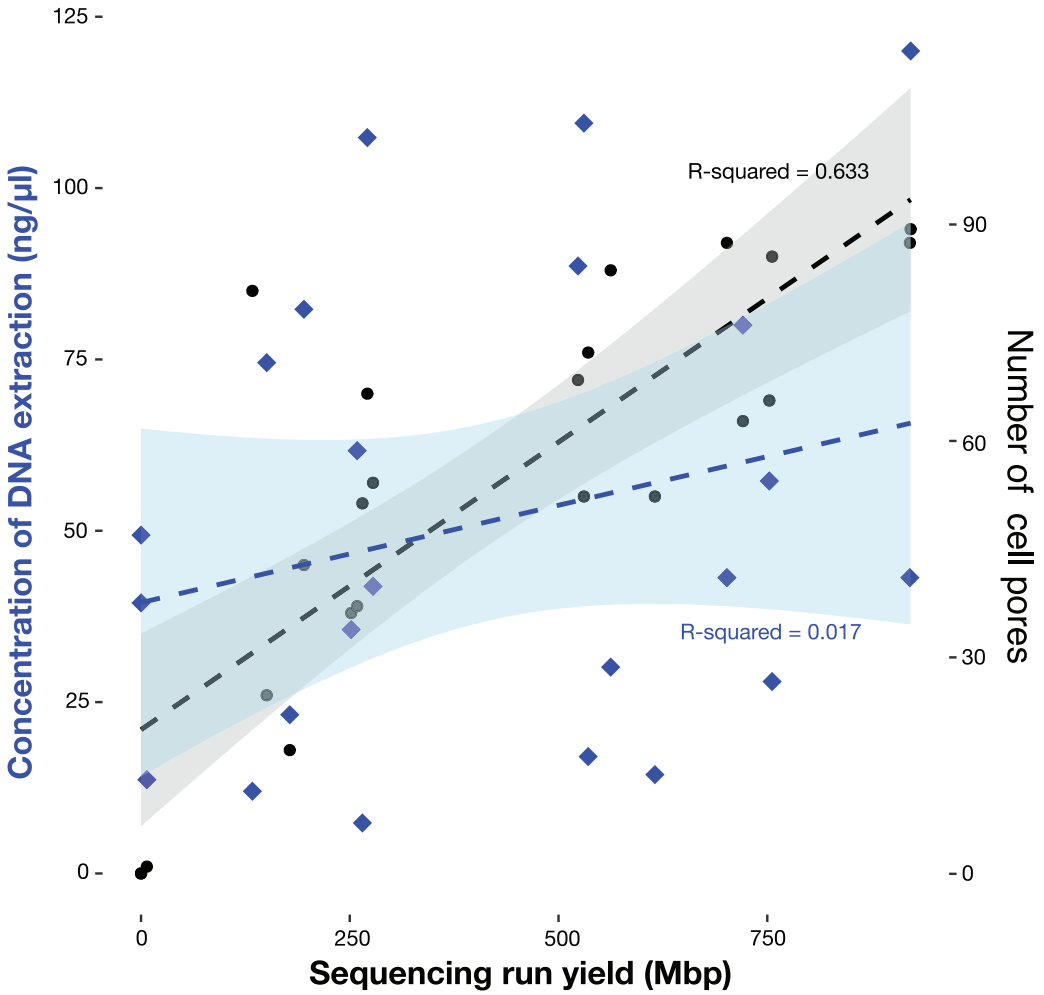

Figure 3 Sequencing output (Mbp) as a function of available pores and concentration of DNA extraction of the 21 individuals sequenced in this study. Available pores are shown in black dots and the concentration of DNA extraction in blue diamonds.

Full-size DOI: 10.7717/peerj.17887/fig-3

value increased compared to the initial concentration of the extracted DNA (Table 1). In most sequencing runs we used Flongle flow cells with more than fifty pores available for immediate sequencing, nonetheless, in three Flongles, the pores were clogged, most likely because the use of cell priming and library loading reagents from the Ligation Sequencing Kit stored in plastic vials, as opposed to the recommended use of reagents stored in the glass vials of the Flongle expansion kit. This means that 21 sequencing runs were successful, while experiments EAFIT-M2126, EAFIT-M2153, and EAFIT-M2154 had insufficient or no sequencing output, and therefore, they were not considered in subsequent analyses. Our sequencing yield varied between 133.35 and 921.72 Mbp and was directly proportional to the number of available pores for immediate sequencing but not to the DNA library concentration (Fig. 3, Table 1). After quality filtering, an average of 82.58% of the reads from all our sequencing runs have a Q score ≥ 7, which represents between 30,019 and 357,353 reads. Finally, the recovered mtDNA

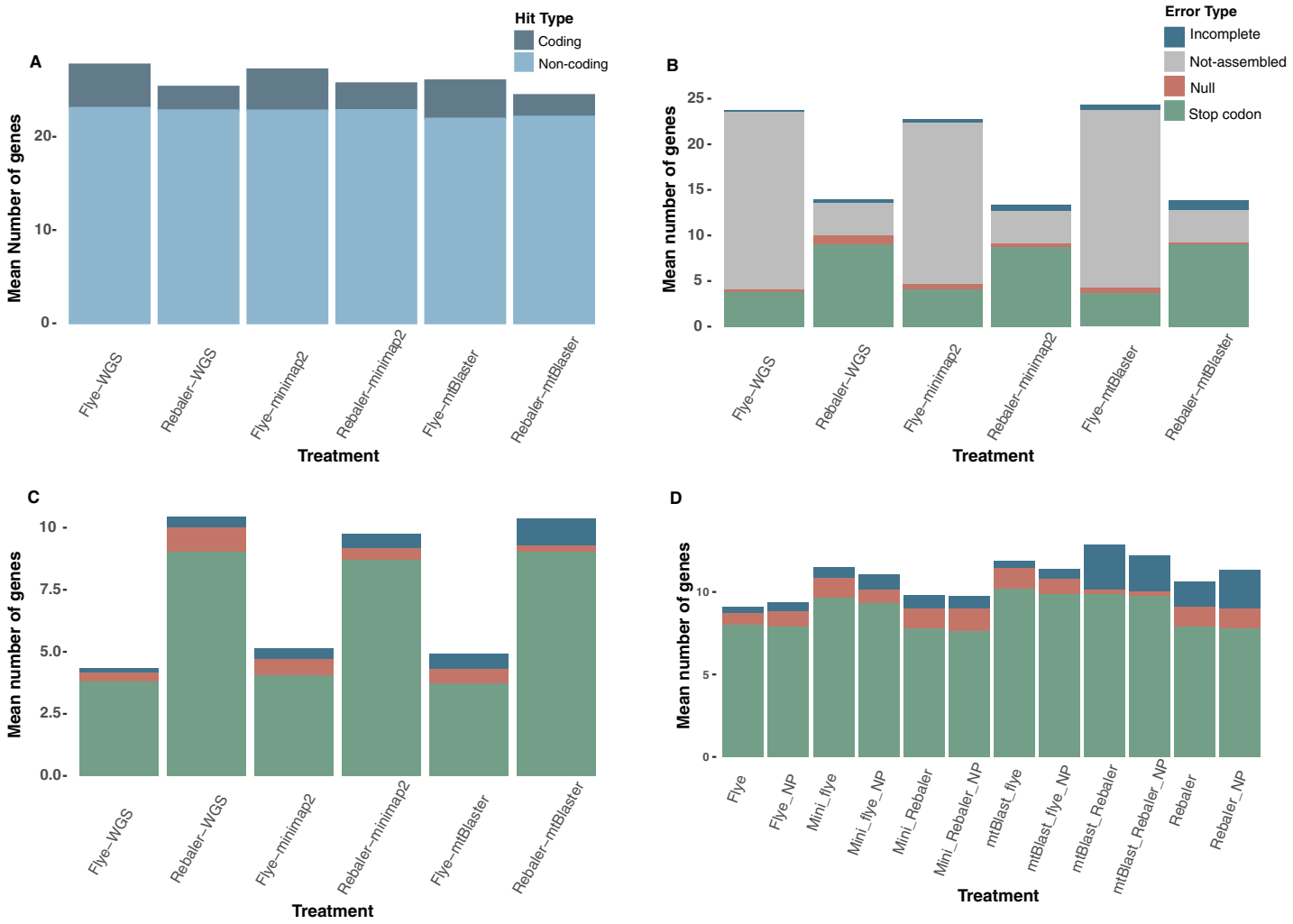

Figure 4 Quality evaluation of mitochondrial DNA reconstruction pipelines. (A) Hit composition of the recovered assemblies, (B) error composition including the non-assembled category, (C) error composition without the non-assembled category, and (D) error composition including polished assemblies with Nanopolish (NP) *versus* non-polished treatments. Full-size DOI: 10.7717/peerj.17887/fig-4

fraction was low, ranging between 0.07% and 0.05%, for Minimap2 and mtBlaster analyses, respectively.

Mitogenome reconstruction

We implemented a total of 252 assembly runs, resulting in 176 successful assemblies (complete and incomplete mitochondrial genomes), from which 100 yielded complete mitochondrial genomes (Table S2). Rebaler (reference-based assembler) recovered a higher number of assemblies regardless of the data type (WGS, Minimap2, mtBlaster) and was able to reconstruct 90.47% of the 21 possible mitogenomes, while Flye (*de novo* assembler) was only able to assemble 50% (Fig. 4, Table 2). On the other hand, Flye assembled the majority of mitogenomes of sequencing runs with more than 550 Mbp, but in sequencing runs with outputs below 550 Mbp Flye was only able to assemble 31.2% of the genomes, regardless of data type (Table 2). In contrast, Rebaler does not seem to be

Table 2 Percentage of mitogenomes assembled by Flye and Rebaler for each data type, according to sequencing yield range.

Assembler	Sequencing yield (Mbp)	% Assembled mitogenomes		
		WGS	Minimap2	mtBlaster
Flye	≥550	77.8	77.8	66.7
	<550	25.0	33.3	33.3
Rebaler	≥550	88.9	100.0	88.9
	<550	91.7	83.3	91.7

Table 3 Assembly quality evaluation based on the assembler and dataset treatment used.

Assembler	WGS	Minimap2	mtBlaster
Assembly hits			
Flye	**8.57%** 27.75	**5.47%** 27.23	**6.06%** 26.05
Rebaler	25.37	25.74	24.47
Assembly errors			
Flye	9.25	9.77	10.95
Rebaler	20.47% 11.63	13.23% 11.26	12.60% 12.53

Note:
Assembly hits (coding genes and non-coding genes) and assembly errors (incomplete, null, and stop codon) excluding non-assembled genes of the mitogenomes assembled by Flye and Rebaler for each data type. The percentage corresponds to the increment of assembly hits (bold) or assembly errors (underline).

affected by the amount of sequencing yield as it assembled 89.6% of mitogenomes, even in runs with less than 550 Mbp regardless of data type (Table 2). Nonetheless, although Rebaler assembles a larger number of mitogenomes, we found that Flye performs better given the lower number of assembly errors (Fig. 4A, Table 3). On average, for all data types, Flye has 6.7% more assembly hits, while Rebaler on average has 15.4% more assembly errors (Table 3, S3). Our error composition comparison shows a higher number of errors in those assemblies obtained from mtBlaster-filtered data (Fig. 4B, Table S3). Overall, assemblies reconstructed with WGS and Minimap2-filtered data have a similar error composition; however, WGS genomes have a better quality than those obtained from a mtDNA filtering processes (mtBlaster or Minimap2) (Fig. 4A, Table S3). If the unassembled genomes are not considered, the mtBlaster+Rebaler pipeline generated the poorest quality assemblies, while those reconstructed using WGS+Flye had the best quality. Therefore, Flye outperformed Rebaler—in terms of assembly quality—in the genomes that were able to reconstruct (Figs. 4A, 4C). Noteworthy, we found that polishing with Nanopolish had a homogeneous effect for all assemblies with no quality improvement, regardless of data type or assembler used (*de novo*, or reference-based) (Fig. 4D, Table S4).

The mitochondrial genome depth (X) exhibited variations depending on the data type and assembler. In general, for WGS+Flye assemblies, the depth ranged from 3X to 17X (Table S2), while for genomes obtained through the WGS+Rebaler pipeline, the average depth ranged from 0.04X to 55.59X (Table S2). It is worth noting that pipelines based on

mitochondrial reads filtering tended to yield higher average depth values than the WGS-based genome assemblies (Table S2). For example, the assemblies obtained through Minimap2+Flye showed depth values ranging from 4.16X to 63.17X, and those from Minimap2+Rebaler had average depths between 2.42X and 140.53X (Table S2). Likewise, genomes obtained from the mtBlaster+Flye pipeline ranged between 3.09X and 57.66X, and mtBlaster+Rebaler between 2.92X and 95.71X (Table S2).

Species identification based on phylogenetic analyses

From our assembled mitochondrial genomes, we were able to obtain 15 Cytb (12 species) and 13 COI (10 species) good-quality sequences susceptible to be included in phylogenetic analyses (Tables 4 and S1). For individuals of the genus *Carollia* (EAFIT-M2157 and EAFIT-M2168) only good-quality Cytb sequences were obtained, and the species (morphologically identified) *Anoura geoffroyi*, *Glossophaga soricina*, *Molossus molossus*, *Sturnira bogotensis* (one individual), and *Zygodontomys brevicauda* lacked sufficient quality for both mitochondrial markers (showed stop codons) and were excluded from the phylogenetic analyses (Table S1). We obtained nine Cytb datasets (alignments) which include 2,578 sequences (File S2) distributed in nine genera (ingroups). The sequence lengths range between 180 and 1,143 bp, resulting in 87.94% nucleotide coverage (Table S1). Similarly, for COI we obtained eight datasets, for eight different genera (ingroups), which include 2,395 sequences (File S2). COI sequences range in length between 201 and 657 bp, representing an average nucleotide coverage of 98.48% (Table S1).

The Cytb and COI topologies are consistent with the literature and show in most cases monophyletic clades currently recognized as species or subspecies (*Gutiérrez, Jansa & Voss, 2010*; *Moratelli et al., 2013*; *Velazco & Patterson, 2013*; *Almendra et al., 2018*; *Giménez, Giannini & Almeida, 2019*; *Rodríguez-Posada et al., 2021*), but clades that exhibit paraphyletic "species" (see *Myotis*) are just the result of the current lack of resolution on their systematics (*Larsen et al., 2012*). Noteworthy, all of the Cytb and COI phylogenetic analyses match the morphological identification (Table 4). For our Cytb analyses, sequences of the species *Akodon affinis, Artibeus lituratus, Carollia brevicauda, Histiotus humboldti, Myotis riparius, Sturnira bogotensis, Marmosa isthmica*, and *Handleyomys alfaroi*, each formed a strongly supported monophyletic group with other sequences of the same species (Fig. 5). The species *Myotis caucensis, Nephelomys childi*, and *N. pectoralis* were not part of monophyletic groups with other conspecifics because there are no available Cytb sequences of these species in the Genbank and consequently they represent new genetic data for these lineages (Fig. 5). For the COI marker, the species *Artibeus lituratus, Myotis caucensis, M. riparius, Sturnira bogotensis, Handleyomys alfaroi, Nephelomys childi*, and *N. pectoralis* are part of haplogroups with other conspecific sequences, while *Histiotus humboldti, Akodon affinis*, and *Marmosa isthmica* represent the first available sequence for each species (Fig. 6).

Table 4 Results of our species identification pipeline at three taxonomic levels (Metazoa, order, family) and our morphology and phylogenetic systematics identifications.

		Metazoa-level					Order-level					Family-level							
Museum number	Number of mt reads	Number of hits	% reads correct Genus	% reads correct species	Metazoa ID	Percentage of Identity	Number of hits	% reads correct Genus	% reads correct species	Order ID	Percentage of Identity	Number of hits	% reads correct Genus	% reads correct species	Family ID	Percentage of Identity	Morphological ID	Phylogenetic ID CytB	Phylogenetic ID COI
EAFIT-M 2126	141	17	64.71	47.06	***Myotis riparius***	93.29	12	91.67	66.67	***Myotis riparius***	93.29	12	100.00	66.67	***Myotis riparius***	93.29	***Myotis riparius***	***Myotis riparius***	***Myotis riparius***
EAFIT-M 2106	187	44	65.91	0.00	*Nephelomys albigularis*[*]	96.39	41	70.73	0.00	*Nephelomys albigularis*[*]	96.39	41	70.73	0.00	*Nephelomys albigularis*[*]	96.39	*Nephelomys pectoralis*	*Nephelomys pectoralis*	*Nephelomys pectoralis*
EAFIT-M 2143	92	20	65.00	5.00	*Histiotus macrotus*	91.80	13	100.00	7.69	*Histiotus macrotus*	91.81	13	100.00	7.69	*Histiotus macrotus*	91.81	*Histiotus humboldti*	*Histiotus humboldti*	*Histiotus humboldti*
EAFIT-M 2107	127	21	47.62	47.62	***Handleyomys alfaroi***	96.20	19	52.63	52.63	***Handleyomys alfaroi***	96.20	19	52.63	52.63	***Handleyomys alfaroi***	96.20	***Handleyomys alfaroi***	***Handleyomys alfaroi***	***Handleyomys alfaroi***
EAFIT-M 2124	365	32	53.13	15.63	*Histiotus macrotus*	96.36	20	85.00	25.00	*Histiotus macrotus*	96.36	21	80.95	23.81	*Histiotus macrotus*	96.36	*Sturnira bogotensis*	Defficient quality sequence	Defficient quality sequence
EAFIT-M 2116	162	18	66.67	55.56	*Carollia perspicillata*[**]	96.53	13	92.31	76.92	*Carollia perspicillata*[**]	96.53	13	92.31	76.92	*Carollia perspicillata*[**]	96.526	*Carollia brevicauda*	*Carollia brevicauda*	Defficient quality sequence
EAFIT-M 2160	547	45	35.56	11.11	*Histiotus macrotus*	93.49	32	56.25	15.63	*Histiotus macrotus*	93.49	34	55.88	17.65	*Histiotus macrotus*	93.49	*Anoura geoffroyi*	Defficient quality sequence	Defficient quality sequence
EAFIT-M 2157	194	17	64.71	64.71	***Artibeus lituratus***	97.54	11	100.00	100.00	***Artibeus lituratus***	97.54	11	100.00	100.00	***Artibeus lituratus***	97.54	***Artibeus lituratus***	***Artibeus lituratus***	***Artibeus lituratus***
EAFIT-M 2164	99	3	33.33	33.33	***Glossophaga soricina***	99.25	1	100.00	100.00	***Glossophaga soricina***	99.25	1	100.00	100.00	***Glossophaga soricina***	99.25	***Glossophaga soricina***	Defficient quality sequence	Defficient quality sequence
EAFIT-M 2158	441	28	53.57	50.00	***Artibeus lituratus***	95.21	17	88.24	82.35	***Artibeus lituratus***	95.21	17	88.24	82.35	***Artibeus lituratus***	95.21	***Artibeus lituratus***	***Artibeus lituratus***	***Artibeus lituratus***
EAFIT-M 2167	195	9	33.33	11.11	***Sturnira bogotensis***	95.66	3	100.00	33.33	***Sturnira bogotensis***	95.66	4	100.00	25.00	***Sturnira bogotensis***	95.66	***Sturnira bogotensis***	***Sturnira bogotensis***	***Sturnira bogotensis***

(Continued)

Table 4 (continued)

		Metazoa-level					Order-level					Family-level							
Museum number	**Number of mt reads**	**Number of hits**	**% reads correct Genus**	**% reads correct species**	**Metazoa ID**	**Percentage of Identity**	**Number of hits**	**% reads correct Genus**	**% reads correct species**	**Order ID**	**Percentage of Identity**	**Number of hits**	**% reads correct Genus**	**% reads correct species**	**Family ID**	**Percentage of Identity**	**Morphological ID**	**Phylogenetic ID CytB**	**Phylogenetic ID COI**
EAFIT-M 2166	323	19	68.42	47.37	*Carollia perspicillata***	97.60	13	100.00	69.23	*Carollia perspicillata***	97.60	13	100.00	69.23	*Carollia perspicillata***	97.60	*Carollia brevicauda*	*Carollia brevicauda*	Defficient quality sequence
EAFIT-M 2165	817	101	21.78	4.95	***Nephelomys childi***	92.66	81	28.40	7.41	***Nephelomys childi***	92.66	86	27.91	6.98	***Nephelomys childi***	92.66	***Nephelomys childi***	***Nephelomys childi***	***Nephelomys childi***
EAFIT-M 2168	287	24	25.00	25.00	*Myotis nigricans***	93.32	8	100.00	75.00	*Myotis nigricans***	93.32	8	100.00	75.00	*Myotis nigricans***	93.32	*Myotis caucensis*	*Myotis caucensis*	*Myotis caucencis*
EAFIT-M 2151	569	71	14.08	9.86	***Nephelomys childi***	93.48	62	32.26	22.58	***Nephelomys childi***	93.70	59	33.90	23.73	***Nephelomys childi***	93.70	***Nephelomys childi***	***Nephelomys childi***	***Nephelomys childi***
EAFIT-M 2159	390	22	36.36	0.00	*Histiotus macrotus*	94.67	22	36.36	0.00	*Histiotus macrotus*	94.67	22	40.91	0.00	*Histiotus macrotus*	94.67	*Molossus molossus*	Defficient quality sequence	Defficient quality sequence
EAFIT-M 2150	788	119	57.98	9.24	*Akodon sp.***	97.09	103	63.11	10.68	*Akodon sp.***	**97.09**	106	53.77	10.38	***Akodon affinis***	95.841	***Akodon affinis***	***Akodon affinis***	***Akodon affinis***
EAFIT-M 2152	157	50	46.00	10.00	***Marmosa robinsoni isthmica***	93.01	49	28.57	14.29	***Marmosa robinsoni isthmica***	93.01	49	28.57	14.29	***Marmosa robinsoni isthmica***	93.01	***Marmosa isthmica***	***Marmosa isthmica***	***Marmosa isthmica***
EAFIT-M 2161	464	136	30.88	2.21	***Marmosa robinsoni***	92.38	155	21.29	3.87	***Marmosa robinsoni isthmica***	92.21	155	21.29	3.87	***Marmosa robinsoni isthmica***	92.21	***Marmosa isthmica***	***Marmosa isthmica***	***Marmosa isthmica***
EAFIT-M 2153	384	35	0.00	0.00	*Histiotus macrotus*	92.28	25	0.00	0.00	*Histiotus macrotus*	92.28	26	0.00	0.00	*Histiotus macrotus*	92.28	*Zygodontomys brevicauda*	Defficient quality sequence	Defficient quality sequence

Notes:

For each dataset, the number of mitochondrial reads is shown. At every level of analysis, the number of total hits recovered using mtBlaster is included. Species with bolded letters (at the three levels of analysis) correspond to identifications that match the other two lines of evidence (*i.e.*, morphology and phylogenetic analyses). Underlined text correspond to specimens for which the morphological and phylogenetic identifications (when available) did not match the one obtained through our species identification pipeline.

* Species lacking mitochondrial sequences in GenBank as of March 2023.

** Species where the most reliable hit corresponds to a misidentified GenBank sequence.

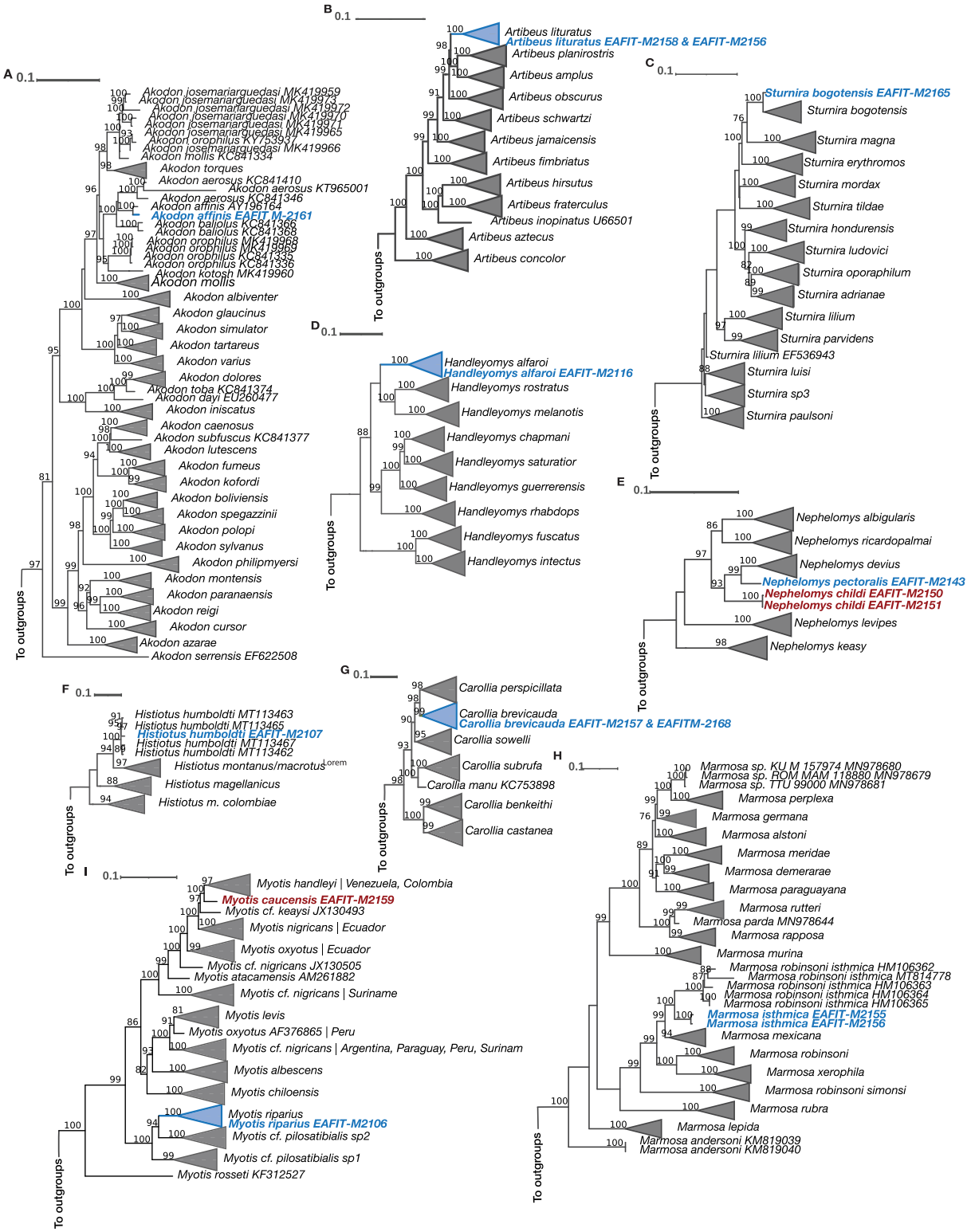

Figure 5 **Maximum Likelihood reconstructions of the Cytb mitochondrial marker for all the genera represented in our taxon sampling.** (A) *Akodon*, (B) *Artibeus*, (C) *Sturnira*, (D) *Handleyomys*, (E) *Nephelomys*, (F) *Histiotus*, (G) *Carollia*, (H) *Marmosa*, (I) *Myotis*. Terminals of non--collapsed clades include GenBank accession numbers. Phylogenies that exhibit paraphyletic species show the corresponding localities of each sequence. Colored terminals correspond to our sequenced material. Bootstrap support values above 75% are shown above branches.

Full-size DOI: 10.7717/peerj.17887/fig-5

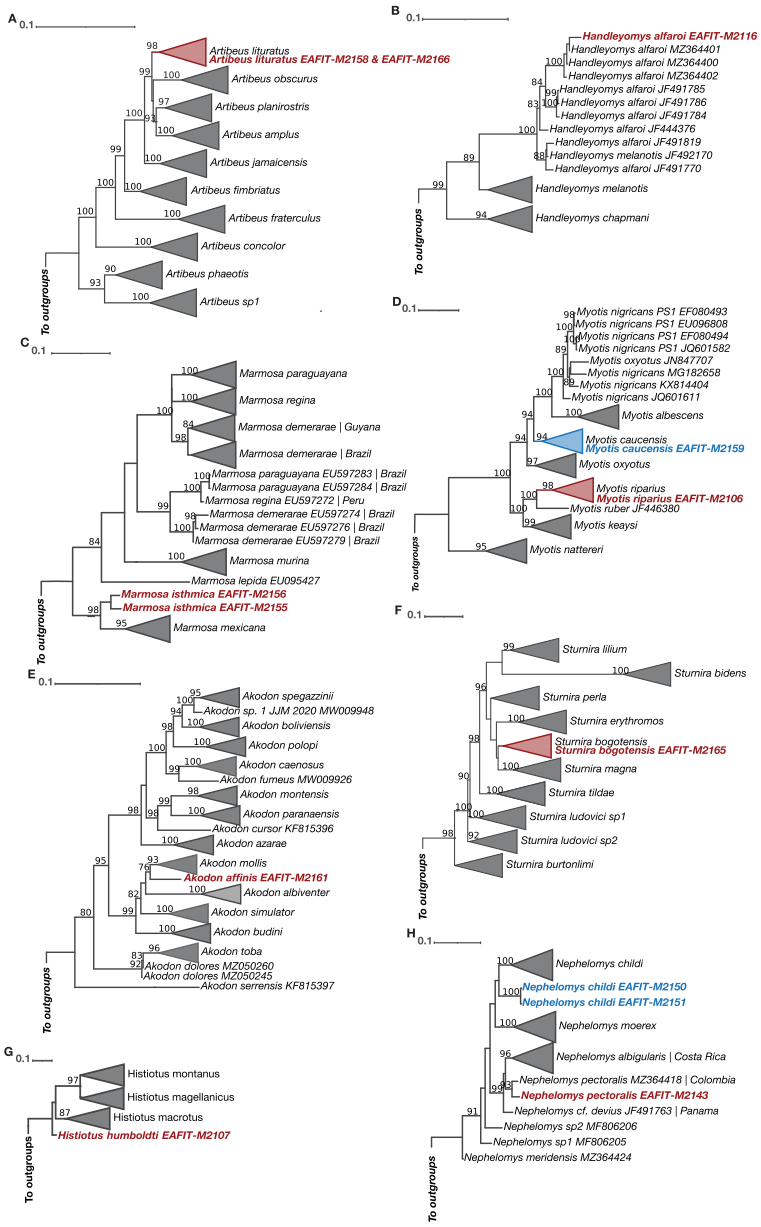

Figure 6 Maximum Likelihood reconstructions of the COI mitochondrial marker for all the genera represented in our taxon sampling. (A) *Artibeus*, (B) *Handleyomys*, (C) *Marmosa*, (D) *Myotis*, (E) *Akodon*, (F) *Sturnira*, (G) *Histiotus*, (H) *Nephelomys*. Terminals of non-collapsed clades include GenBank accession numbers. Phylogenies that exhibit paraphyletic species show the corresponding localities of each sequence. Colored terminals correspond to our sequenced material. Bootstrap support values above 75% are shown above branches. Full-size DOI: 10.7717/peerj.17887/fig-6

Species identification pipeline with ONT raw reads

In 20 of the analyzed datasets, it was possible to recover mitochondrial reads, a single dataset was excluded from subsequent analyses (EAFIT-M2124 corresponding to the species *Rhipidomys latimanus*) because mtBlaster was unable to recover mitochondrial reads (Table S2). Our species identification pipeline did not show any relevant difference

between the three taxonomic levels of analyses with respect to the "top" (most reliable) identification (Table 4). In all three analyses (Metazoa, order, family), for 10 samples, our pipeline matched the identification of the morphological+phylogenetic analyses (species *Artibeus lituratus, Glossophaga soricina, Myotis riparius, Sturnira bogotensis, Handleyomys alfaroi, Nephelomys childi, Marmosa isthmica*), four other samples were recovered by our pipeline with the correct genus but the species was incorrect due to misidentification of the sequence in the DNA library (species *Carollia brevicauda, Myotis caucensis, Akodon affinis, Nephelomys pectoralis*), and for one sample we also recovered the correct genus but the species was incorrect because of the lack of sequences of such species in the mtDNA library (*Nephelomys pectoralis*) (Table 4). Conversely, for five samples our pipeline recovered an incorrect identification even though the mtDNA library contains sequences of each species; however, from those four samples, only one recovered the incorrect genus and species (*Handleyomys rhabdops* instead of the correct *Zygodontomys brevicauda*), but the remaining four were allocated in the correct genus (*Anoura, Histiotus, Molossus, Sturnira*) (Table 4).

We also evaluated the effectiveness of the species identification pipeline for the genus and species by calculating the percentage of reads corresponding to the correct genus and species for each dataset at all levels of analyses (Metazoa, order, family). For our three levels of analyses, only three samples did not show any reads associated with the correct identification of the species; however, one of those datasets did not contain mitochondrial sequences for the species (*Nephelomys pectoralis*) whereas the other two datasets were unable to associate the reads with its correct species (*Molossus molossus, Zygodontomys brevicauda*) (Table 4, Fig. 7). For the other remaining 17 samples, all analyses—at all three levels—included reads associated with the correct species (Table 4). Overall, the analyses showed a consistent pattern, they were more effective at providing the correct identification as more taxonomic information was available, that is, a higher number of correct reads for the order and family databases than for the Metazoa database (Table 4, Fig. 7). We found that 10 of the order and family datasets provided more than 80% of the reads with the correct genus identification and seven datasets with at least 65% of their reads associated with the correct species (Fig. 7). There seems to be an additional pattern, the more sequences of the correct genus and species in the databases, the greater the number of reads associated with the correct species (Fig. 7). Generally, mitochondrial datasets (order and family) with more than 812 sequences provided >80.09% of correct identifications at the genus level, and datasets larger than 97 sequences produced >66.67% of correct identifications at the species level (Fig. 7).

We observed variations in the processing time among the three taxonomic levels of analyses. When conducting a BLAST search without any prior sample identification (*i.e.*, Metazoa), the analysis took between 01:26:37 and 01:38:57 h for datasets with the smallest and highest number of potential mitochondrial reads respectively (92 and 817 reads). In contrast, the analysis against the order-level dataset required only between 00:00:08 s and 00:01:28 min, while the family-level analysis took between 00:00:05 to 00:00:32 s. These findings highlight the substantial differences in computational time required depending on

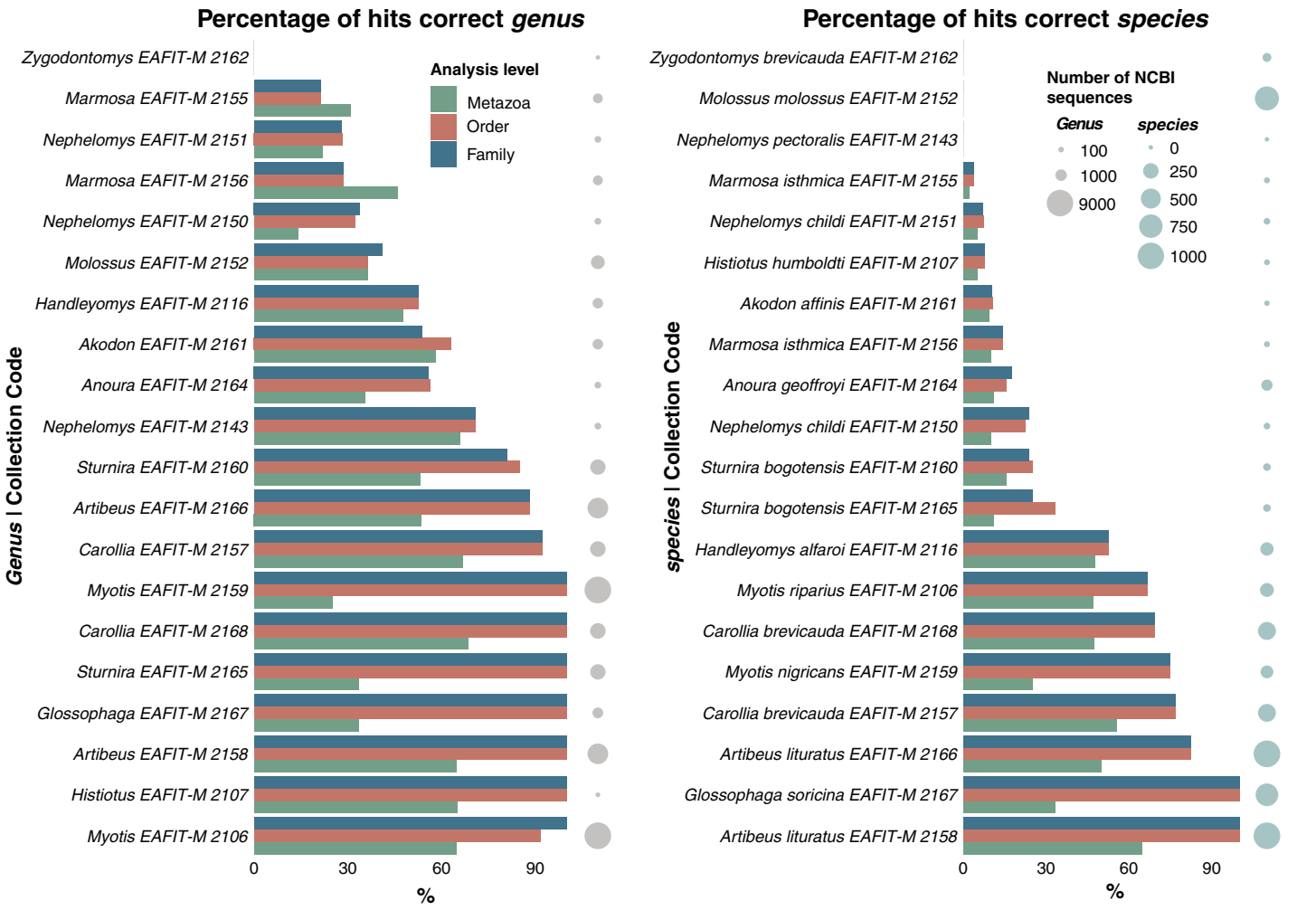

Figure 7 Accuracy comparison of our species identification pipeline based exclusively on ONT reads, using three different taxon-level databases (Metazoa, order, family). Percentage of reads that match the correct identification of the genus (left) and species (right) based on our morphological and phylogenetic systematics identification. The size of the dots indicate the number of available NCBI sequences for each genus (left) and species (right).

Full-size DOI: 10.7717/peerj.17887/fig-7

the database size and the relevance of additional taxonomic information when performing these identifications.

DISCUSSION

Mitogenome reconstruction

In this study, we evaluated various computational routines for reconstructing mitochondrial genomes from non-model organisms, exclusively using ONT reads and found that this sequencing technology still exhibits limitations in comparison to other technologies. It maintains a higher raw sequencing error rate, approximately 5% (with the now deprecated R9 chemistry), which is notably higher than the approximately 0.3% error rate seen in Illumina (*Baeza & García-De León, 2022*). These errors are mainly attributed to mismatches between the raw signal and the basecalling process, generating modified

nucleotides and homopolymer sequences (*Magi, Giusti & Tattini, 2017*). To address this issue, progress has been made in basecalling algorithms such as Guppy, improving the accuracy of basecalling compared to the older Albacore software (*Baeza, 2020*; *Baeza & García-De León, 2022*). Another strategy is based on error correction by implementing polishing programs specialized to deal with noisy long reads (*Filipović et al., 2022*). However, consistent with previous findings (*Baeza & García-De León, 2022*), we did not observe any noticeable improvement in the quality of the assemblies after applying a round of polishing with Nanopolish (Fig. 3).

The error rate limitations of ONT can also be effectively addressed through strategies such as maintaining a high read depth (*Pollard et al., 2018*; *Ramón-Laca, Gallego & Nichols, 2023*). Studies using PCR-free genome skimming approach with long reads, to reconstruct complete mitochondrial genomes have reported varying average genome depths, between 9X and 78X (*Malukiewicz et al., 2021*; *De Vivo et al., 2022*; *Franco-Sierra & Díaz-Nieto, 2020*). In our study, although most genomes assembled with either Flye or Rebaler exhibited depth values that fell within the range observed in the previously mentioned studies, only three routines produced high quality and error-free genomes: *Nephelomys childi* EAFIT-M2150 (17.00X), EAFIT-M2151 (36.62X), and *Akodon affinis* EAFIT-M2161 (26.09X), with NCBI accession numbers PP556414, PP555285 and PP556413 respectively (Table S2). We also observed that mitochondrial filtering increased depth values regardless of the assembler or filtering pipeline used (Table S2). This could be a signal that when the primary objective is to obtain good quality mitogenomes, filtering out the mitochondrial reads can increase the depth and thus the quality of the recovered genome (*Ramón-Laca, Gallego & Nichols, 2023*). Although depth is a key factor for reconstructing good quality genomes, our species identification pipeline does not heavily depend on this variable. Many datasets with shallow depths provided accurate identifications. For instance, the highest depth for the assembly of dataset EAFIT-M2158 was 4.72X (mtBlaster+Rebaler); however, this dataset produced correct identifications at all three levels of analyses (see below). However, although mtDNA filtering can increase genome depth, our analyses show that filtering with Minimap2 or mtBlaster does not necessarily increase the number of successful assemblies, in fact, we observe an increase in the number of errors such as incomplete genes, null genes, and premature stop codons (Fig. 4). This phenomenon may occur because the implemented algorithms fail to obtain all the mitochondrial reads or might recover erroneous mitochondrial sequences, such as NUMTs.

As for the assemblers, Flye is more accurate and has better quality overall; however, it requires higher sequencing throughput (≥550 Mpb) to assemble genomes, compared to Rebaler which, although it is able to assemble most mitogenomes, even from datasets with limited sequencing yield (<550 Mbp), it exhibits higher error rate (Table 2). Indeed, recent studies that have conducted comparisons of the resulting mitogenomes using these two assemblers have found similar results (*Baeza, 2020*; *Baeza & García-De León, 2022*). This phenomenon may have two different explanations. On the one hand, Rebaler uses Minimap2 for mapping potential mitochondrial reads to reference genomes, which has

been shown to be difficult and leading to a decrease in the quality of resulting assemblies (*Baeza, 2020*; *Baeza & García-De León, 2022*). On the other hand, Flye implements a graph-based approach for single-molecule sequencing data that enhances genome assembly quality of noisy long-reads, like the ones generated by ONT (*Kolmogorov et al., 2019*; *Zhang et al., 2022*). Additionally, the option for metabarcoding assemblies included in Flye seems to work well in genomes that have low or uneven coverage (*Kolmogorov et al., 2019*), which is the case of mitochondrial genomes from whole genome shotgun sequencing.

Species identification pipeline with ONT raw reads

The implemented methodology exhibits promising results in the field of species identifications without the need of industrious laboratory work (no PCR), or expert knowledge on taxonomy and phylogenetic systematics; however, it will still need to optimize two main aspects before it can be extensively implemented for species identification. First, our methodology employs a Flongle flow cell (non-reusable; 100 USD) for sequencing, which yields up to 900 Mb of data and enables the identification of a single individual. However, the mtDNA fraction for species identification constitutes only 0.07% of the total sequencing output, resulting in most of the economic expenditure being allocated to sequencing data not used for identification purposes. Second, our methodological approach is not yet real-time, as it needs a minimum of 25 h to achieve species identification (24 h of sequencing plus <1 h for the species identification pipeline). Both problems can be addressed by using the standard and reusable Flow Cell (up to 50 Gb). This Flow Cell enables sequencing of the necessary output for individual species identification. It can be washed and reused for subsequent sequencing experiments, reducing sequencing costs per individual. In cases where time constraints are significant, it may be preferable to choose a Flow Cell that can produce the minimum sequencing output required for identification in a shorter time frame.

Our pipeline was able to correctly identify most (95%) of the samples to the genus level and half of the samples (50%) to the species level, but this last value can be up to 75% if we correct the identification of the sequences from BOLD or NCBI (Table 4). We found that species identification effectiveness was partially dependent on the sequencing output (Fig. 3). Overall, datasets that exceeded 550 Mbp, had more effective identifications compared to those that were below this sequencing yield. In some cases, not only the amount of data but the quality of the reads was related to the inability of our pipeline to accurately assess the taxon identity. In fact, in datasets containing low-quality reads, we encountered challenges not only in obtaining effective identifications, but also in extracting mitochondrial markers with sufficient quality for phylogenetic reconstructions as was the case for datasets EAFIT-M2160, EAFIT-M2164, EAFIT-M2152, and EAFIT-M2162 (Table 4).

Accuracy of identifications at the species and genus-level seems to be dependent on the availability of sequences in DNA repositories (GenBank, BOLD). Our results show that as more mitochondrial DNA data is available, there is a noticeable increase in the frequency

of successful matches (blasted reads) accurately identifying the correct genus and species (Fig. 7). For instance, at the three levels of analysis (Metazoa, order, family) our species identification pipeline identified our *N. pectoralis* individual (dataset EAFIT-M2143) as *N. albigularis*, which can be explained by two facts. First, as of March 2023, there were no *N. pectoralis* sequences available in GenBank or BOLD. Second, *N. pectoralis* is an endemic species from Colombia, recently split from *N. albigularis* which is currently recognized as a species restricted to Peru and Ecuador (*Weksler, Percequillo & Voss, 2006*), thus, it makes sense that the most reliable identification is from a closely related lineage.

Meanwhile, automated species identification methods, including our approach, depend heavily on the identification—scientific name—assigned to sequences from repositories like BOLD and GenBank. Consequently, misidentifications or outdated names in the sequences from DNA databases will obviously produce incorrect identifications. Three cases in our analyses illustrate this scenario. Our two individuals morphologically identified as *Carollia brevicauda*, were defined by our species identification pipeline as *C. perspicillata*. Historically, these two species have been frequently misidentified due to their morphological similarities (*Ruelas & Pacheco, 2022*). However, we were able to obtain high-quality mitochondrial markers for both datasets and their analyses with other congeneric sequences confirmed our initial morphological identification. Moreover, while the "top" identification retrieved through the species identification pipeline was *C. perspicillata*, most of the recovered hits corresponded to *C. brevicauda* (File S1). Therefore, although we were unable to examine the vouchers, we are inclined to believe that the specimen from GenBank corresponding to the most reliable hit is an example of misidentification.

A similar example occurs with the individual EAFIT-M2159, morphologically identified as *Myotis caucensis*, which was identified by our species identification pipeline as *M. nigricans*. Our morphological identification was based on the original description of the species (*Moratelli et al., 2013*). We found that although dorsal and ventral coloration patterns follow more closely the description of *M. handleyi*, dental characters are more closely related to *M. caucensis* (*Moratelli et al., 2013*). In particular, the upper second premolar which is evidently visible in lateral view on the latter species, and is also present in our material (*Moratelli et al., 2013*). Additionally, *M. handleyi* is a recently described species believed to be sister to *M. caucensis* (*Moratelli et al., 2017*), consequently, molecular data uploaded to databases before these taxonomic updates may not accurately reflect the current taxonomy. Sequences previously labeled as *M. nigricans* could pertain to either *M. handleyi* or *M. caucensis*. In fact, our phylogenetic analysis of Cytb supports this hypothesis, as our sequence is sister to *M. handleyi* (Fig. 5). In conclusion, the absence of updated taxonomy associated with sequences poses challenges for the use of genetic data within automated species identification pipelines, such as the one presented here.

The third example corresponds to an individual of the species *Akodon affinis* (EAFIT-M2161). Using the Metazoa and order databases, our sample could not be identified to the species level because the most reliable hit in both cases corresponds to the name *Akodon* sp. (NCBI:txid2964413). Two lines of evidence lead us to believe that this last specimen

corresponds to the species *A. affinis*. First, the specimen was collected in Cauca (Colombia), a locality reported to be within the distribution range of *A. affinis* (*Patton, Pardiñas & D'Elía, 2015*). Second, our phylogenetic analysis supported our morphological identification, as our sequence formed a highly supported monophyletic group with another *A. affinis* sequence (Fig. 5). As in the two previous examples, misidentification decreases the effectiveness of our species identification pipeline.

The mtDNA "*in silico* enrichment" strategy which uses DC-BLAST to isolate mitochondrial reads from WGS sequencing data, proposed by *Franco-Sierra & Díaz-Nieto (2020)* and optimized here, was highly efficient in reducing computational time and increasing the effectiveness (accuracy) of species-level identifications. Consequently, the availability of multiple mitochondrial markers for diverse metazoan organisms allows to simplify species identification process by only using the mitochondrial fraction of the sequenced genome (*Hebert, Ratnasingham & de Waard, 2003*; *Stoekle, 2003*; *Page & Hughes, 2010*; *Armani et al., 2017*; *Elyasigorji et al., 2023*). Noteworthy, our results indicate that the computational pipeline for species identification can be more efficient, but only slightly more accurate, depending on the size of the database implemented for DC-BLAST (Metazoa, order, family). This is because a smaller reference database reduces the amount of computational load, but the gain in accuracy was only reduced to one (see *Akodon affinis*) out of 20 analyses.

Furthermore, automated species identification methods such as the one used in this study should always undergo a critical review process to analyze their results. Although an e-value very close to zero indicates a high level of confidence in a particular identification (*Kerfeld & Scott, 2011*; *Shah et al., 2019*), it is important to consider other output parameters. For example, it is important to account for the read length, as short reads can produce very low e-values because the shorter the sequence the more likely it is to match the nucleotides of a reference sequence (*Kerfeld & Scott, 2011*; *Shah et al., 2019*). Although short reads (ca. 100 bp) can be useful for identification purposes in a phylogenetic scenario (*Díaz-Nieto, Jansa & Voss, 2016*; *Teta & Díaz-Nieto, 2019*), their usefulness seems to be reduced with alignment search algorithms. It is therefore important to examine the proposed identification considering basic natural history information such as distribution and habitat, to determine the validity and meaningfulness of the recovered identification, as explained with the *Nephelomys pectoralis* case.

It is clear that our species identification pipeline is not infallible, and it will likely be benefited from the new ONT chemistry (Ligation Sequencing Kits SQK-LSK114 and Flongle Flowcells R10) as it significantly increases the quality of the sequenced reads from ~90% to ~99% (*Sanderson et al., 2023*; *Ni et al., 2023*). However, to further improve the effectiveness of our pipeline it will be important to increase the recovered mitochondrial fraction, which although concordant with other studies (*Franco-Sierra & Díaz-Nieto, 2020*; *De Vivo et al., 2022*) it is still very small (between 0.01% and 0.06% of total bases). Multiple methodologies have been developed to increase the mtDNA fraction in a sample, such as the use of centrifugation protocols (*Stockburger et al., 2016*; *Gaudioso et al., 2019*; *Ramón-Laca, Gallego & Nichols, 2023*), PCR-based methodologies

(*Zascavage et al., 2019*) and Cas9 approaches (*Ramón-Laca, Gallego & Nichols, 2023*). Nonetheless, additional laboratory steps include specialized equipment (ultracentrifuge) or enzymes whose life depends on constant refrigeration, hinder the implementation of this sequencing technology under field conditions, which we consider is the ultimate goal with this technology. Also, incorporating more reagents and equipment would increase operating costs, reducing its accessibility for institutions with limited budgets. Although, the overall cost of mitochondrial genome sequencing decreases with the implementation of mitochondrial DNA enrichment or sample multiplexing (*Zascavage et al., 2019*; *Ramón-Laca, Gallego & Nichols, 2023*), such approaches require expensive materials and reagents that will intrinsically inflate the protocol costs. Finally, increasing the mitochondrial fraction may provide more accurate identifications, but also augment sample processing time in up to 48 h depending on the protocol (*Zascavage et al., 2019*; *Ramón-Laca, Gallego & Nichols, 2023*), reducing the "real-time" advantage of this approach.

Species identification based on phylogenetic analyses

Our results make an important contribution to the knowledge of small mammal diversity, but also help to fill the gaps that still exist in genomic databases. This project is generating the first complete mitochondrial genomes for the species *Akodon affinis* and *Nephelomys childi*. Additionally, the first COI sequences for five species (*Akodon affinis*, *Sturnira bogotensis*, *Nephelomys pectoralis*, *Histiotus humboldti and Myotis caucensis*); first Cytb sequences for the species *Nephelomys childi*, *N. pectoralis*, and *Myotis caucensis;* the first two (2) sequences for *Nephelomys childi* representing populations from Cordillera Central (Colombia), and the second Cytb sequence for the species *Akodon affinis*.

Our Cytb sequences for *Marmosa isthmica* form a highly supported group sister to the known *M. isthmica* haplogroup with an uncorrected p-distance of 0.068 between clades (Fig. 5). Interspecific distances between congeneric species are similar and even smaller, as is the case of *M. demerarae* and *M. merida* or *M. parda* and *M. rutteri* with p-distance values of 0.07 and 0.05, respectively (*Voss et al., 2020*). *Marmosa isthmica* is distributed from western Panama along the Pacific coast of Colombia to Santa Rosa (Ecuador), including the lowlands of northwestern Colombia, and the intern-Andean valleys of Cauca and Magdalena rivers. Despite extensive revisions (*Rossi, Voss & Lunde, 2010*; *Gutiérrez, Jansa & Voss, 2010*; *Gutiérrez et al., 2014*; *Voss et al., 2020*), most studies have not included genetic information from Colombian localities (*Gutiérrez, Jansa & Voss, 2010*; *Gutiérrez et al., 2014*; *Giarla & Voss, 2020*). Noteworthy, our two *M. isthmica* individuals were collected in dry forest areas of the Cauca River, below 700 masl, whereas most collection sites reported for this species correspond to moist lowland and premontane forests below 1,700 m elevation (*Rossi, Voss & Lunde, 2010*). Consequently, it is important to evaluate whether the observed genetic distance between these two clades suggests the presence of cryptic diversity or is merely the result of an isolation-by-distance effect that will be diluted when an expanded geographical sampling becomes available.

Nephelomys childi is endemic to cloud forest habitats in the three Andean Cordilleras of Colombia (*Patton, Pardiñas & D'Elía, 2015*) and inter-Andean valleys have been proposed

as barriers that prevent migration between populations resulting in the observed pattern of allopatric distributions in the three Cordilleras (*Cárdenas González, 2017*). We found that our COI sequences of *N. childi* form a highly supported clade sister to a haplogroup from the Cordillera Oriental (Fig. 7). We also observed a considerable variation (5.61% in uncorrected p-distance) between these two haplogroups (Fig. 7). Considering these facts, we may be facing two different scenarios. First, there is hidden diversity within the *N. childi* group, and the allopatric distributions reported for the species may have triggered allopatric speciation processes. In fact, this is further substantiated by recent systematic studies on this genus (*Weksler, Percequillo & Voss, 2006*), which have emphasized the need for a taxonomic revision to elucidate the internal phylogenetic relationships and possible presence of cryptic species, establish diagnosis, and clarify the geographical distribution of the species (*Cárdenas González, 2017*; *Ruelas et al., 2021*). In a second scenario, our sequence divergence is just the result of the geographic distance found between populations (*Wright, 1943*; *Meirmans, 2012*). The species *N. childi* is underrepresented in genomic databases. To date, the genus has 169 mitochondrial sequences available on GenBank, of which only five belong to *N. childi* (Cordillera Oriental) and are restricted to the marker COI. Incrementing the molecular sampling along the entire distribution of the species would close the gap in the genetic distance between populations of the species.

CONCLUSIONS

In this study, we explored the use of the ONT MinION device with the Flongle flow cells as a tool for the reconstruction of mitochondrial genomes of small mammals, and to obtain species identifications of non-model mammals from multiple lineages (opossums, rodents, bats) using solely ONT reads. Specifically, we learned that for sequencing runs above 550 Mbp, the mitochondrial genome reconstruction should be performed using pipelines based on *de novo* assemblies, which have a better quality regardless of data type. However, if sequencing runs have less than 550 Mb a better result can be obtained by using a referenced-based assembler. Furthermore, although mitochondrial DNA filtering from WGS data seems to increase the depth of recovered mitogenomes, it does not appear to increase the number of successful assemblies. Instead, mitochondrial DNA filtering produces more assembly errors compared to those obtained from WGS data. Also, our findings suggest that the polishing step is not required as it does not improve the quality of assemblies, instead it is time- and computational-resource consuming.

Although our pipeline has proven not to be flawless, we observed that it produces identifications with more than 95% accuracy in non-model species of Neotropical small mammals. This can become a powerful approach for the description and characterization of the planet's diversity because it can be implemented by any researcher (even without extensive taxonomic, phylogenetic, or morphological training), in real-time and even under field conditions. Species identification results are highly dependent on the availability of data in public repositories and the curatorial (including taxonomy updates) of that information. We found that the more sequences of a taxonomic group are available in a database, the more effective is the identification of an individual with unknown

identity. This demonstrates the importance of contributing to the growing field of species identification and description and closing the gap of knowledge on non-model organism diversity. Accordingly, the results obtained in this study make an important contribution to increasing the knowledge of the genetic diversity of small mammals by making available three new mitochondrial genomes and eight new DNA markers for science.

The findings of this work add to the growing body of data demonstrating the usefulness of ONT as a PCR-free option for sequencing full organellar genomes and obtaining species identifications in a short amount of time and on a wider taxonomic scale. Furthermore, the proposed methodology in this work stands as a relevant tool to help increase the available data for non-model organisms, and possibly it can be applied for onsite non-model species identification.

ACKNOWLEDGEMENTS

We acknowledge the Apolo Scientific Computational Center from EAFIT University and especially Laura Sánchez Córdoba for her constant support and providing access to generous computational resources. We thank Hacienda La Sierra, Universidad EAFIT, and Finca Socolao for their generosity facilitating our logistics during the field work. We thank Esteban Velásquez for his bioinformatic support during the initial stages of the development of this project, and Oscar S. Alzate, Juan M. Lozano, Luis J. Pérez, and Mariana Cruz from the Biodiversity research group for their help with fieldwork activities. We would also like to thank Y. Xilena Rueda for her assistance at the biological collection of Universidad EAFIT.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was funded by Universidad EAFIT through the research grant "Identificación molecular de especies en tiempo real por medio del uso de técnicas de secuenciación de tercera generación" awarded to Juan F. Díaz-Nieto. Additional funding was given to Sara Velasquez-Restrepo by Fundación Alejandro Angel Escobar, Colombia Biodiversa grant (Cohort 2022-2). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Universidad EAFIT.
Fundación Alejandro Angel Escobar, Colombia Biodiversa: Cohort 2022-2.

Competing Interests

Nicolás D. Franco-Sierra is employed by Syndesis Health and is also an ad honorem member of VEDAS Corporación de Investigación e Innovación (VEDAS CII), a non-profit research organization.

Author Contributions

- Sara Velasquez-Restrepo conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Mariana Corrales Orozco conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Nicolás D. Franco-Sierra conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Juan M. Martínez-Cerón analyzed the data, prepared figures and/or tables, and approved the final draft.
- Juan F. Díaz-Nieto conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Animal Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

The Ethichs committee of The Ethics committee of the Universidad EAFIT provided full approval for this research in the November 16th 2021 virtual session. Attached as Supplemental File is the ethics committee approval letter.

Field Study Permissions

The following information was supplied relating to field study approvals (*i.e.*, approving body and any reference numbers):

Field experiments were approved by the "Autoridad Nacional de Licencias Ambientales (ANLA) under resolution 1160 of June 7, 2023".

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The COI and CytB sequences are available at GenBank: OR873409-OR873422; OR988047-OR988062.

The complete genomes are available at GenBank: PP556413, PP556414 and PP555285.

The genomes are also available in the Supplemental Files.

Data Availability

The following information was supplied regarding data availability:

The raw fastq sequences are available at FigShare: Velasquez, Sara; Díaz-Nieto, Juan F. (2024). Quality filtered data. figshare. Dataset. https://doi.org/10.6084/m9.figshare.24669249.v1.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.17887#supplemental-information.

REFERENCES

Almendra AL, González-Cózatl FX, Engstrom MD, Rogers DS. 2018. Evolutionary relationships and climatic niche evolution in the genus *Handleyomys* (Sigmodontinae: Oryzomyini). *Molecular Phylogenetics and Evolution* **128**:12–25 DOI 10.1016/j.ympev.2018.06.018.

American Veterinary Medical Association (AVMA). 2020. AVMA guidelines for the Euthanasia of animals: 2020 edition. *Available at https://www.avma.org/sites/default/files/2020-02/Guidelines-on-Euthanasia-2020.pdf* (accessed 3 December 2023).

Antil S, Abraham JS, Sripoorna S, Maurya S, Dagar J, Makhija S, Bhagat P, Gupta R, Sood U, Lal R, Toteja R. 2022. DNA barcoding, an effective tool for species identification: a review. *Molecular Biology Reports* **50**:761–775 DOI 10.1007/S11033-022-08015-7.

Antonelli A, Sanmartín I. 2011. Why are there so many plant species in the Neotropics? *Taxon* **60(2)**:403–414 DOI 10.1002/tax.602010.

Armani A, Tinacci L, Lorenzetti R, Benvenuti A, Susini F, Gasperetti L, Ricci E, Guarducci M, Guidi A. 2017. Is raw better? A multiple DNA barcoding approach (full and mini) based on mitochondrial and nuclear markers reveals low rates of misdescription in sushi products sold on the Italian market. *Food Control* **79(5)**:126–133 DOI 10.1016/J.FOODCONT.2017.03.030.

Baeza JA. 2020. Yes, we can use it: a formal test on the accuracy of low-pass nanopore long-read sequencing for mitophylogenomics and barcoding research using the Caribbean spiny lobster *Panulirus argus*. *BMC Genomics* **21**:103 DOI 10.1186/s12864-020-07292-5.

Baeza JA, García-De León FJ. 2022. Are we there yet? Benchmarking low-coverage nanopore long-read sequencing for the assembling of mitochondrial genomes using the vulnerable silky shark *Carcharhinus falciformis*. *BMC Genomics* **23**:352 DOI 10.1186/S12864-022-08482-Z.

Benítez-Páez A, Portune KJ, Sanz Y. 2016. Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION™ portable nanopore sequencer. *Gigascience* **5(1)**:s13742-0162 DOI 10.1186/s13742-016-0111-z.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* **22(3)**:148–155 DOI 10.1016/j.tree.2006.11.004.

Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **76(4)**:1967–1971 DOI 10.1073/PNAS.76.4.1967.

Byrne AQ. 2023. Reimagining the future of natural history museums with compassionate collection. *PLOS Biology* **21(5)**:e3002101 DOI 10.1371/journal.pbio.3002101.

Cárdenas González C. 2017. Taxonomía y distribución del género de roedores andinos: *Nephelomys* (Cricetidae: Sigmodontinae) en Colombia. MSc, thesis, Universidad Nacional de Colombia, Bogotá, Columbia.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**:403 DOI 10.1186/1471-2105-10-421.

Cicero C, Mason NA, Jiménez RA, Wait DR, Wang-Claypool CY, Bowie RCK. 2021. Integrative taxonomy and geographic sampling underlie successful species delimitation. *Ornithology* **138(2)**:28 DOI 10.1093/ornithology/ukab009.

Ciocca DR, Delgado G. 2017. The reality of scientific research in Latin America; an insider's perspective. *Cell Stress and Chaperones* **22(6)**:847–852 DOI 10.1007/s12192-017-0815-8.

Clarke AC, Prost S, Stanton JAL, White WTJ, Kaplan ME, Matisoo-Smith EA, Adhikarla S, Adler CJ, Balanovska E, Balanovsky O, Bertranpetit J, Comas D, Cooper A,

der Sarkissian CSI, Dulik MC, Gaieski JB, GaneshPrasad AK, Haak W, Haber M, Li LJ, Martínez-Cruz B, Merchant NC, Mitchell RJ, Owings AC, Parida L, Pitchappan R, Platt DE, Quintana-Murci L, Renfrew C, Lacerda DR, Royyuru AK, Santos FR, Schurr TG, Soodyall H, Hernanz S, Swamikrishnan P, Tyler-Smith C, Santhakumari AV, Vieira PP, Vilar MG, Zalloua PA, Ziegle JS, Wells RS. 2014. From cheek swabs to consensus sequences: an A to Z protocol for high-throughput DNA sequencing of complete human mitochondrial genomes. *BMC Genomics* **15(1)**:1–12 DOI 10.1186/1471-2164-15-68.

Cleman N, Rowe KMC, Rowe KC, Raadik T, Gomon M, Menkhorst P, Sumner J, Bray D, Norman M, Melville J. 2014. Value and impacts of collecting vertebrate voucher specimens, with guidelines for ethical collection. *Memoirs of Museum Victoria* **72**:141–151 DOI 10.24199/j.mmv.2014.72.09.

Costello MJ, Beard KH, Corlett RT, Cumming GS, Devictor V, Loyola R, Maas B, Miller-Rushing AJ, Pakeman R, Primack RB. 2016. Field work ethics in biological research. *Biological Conservation* **203(8)**:268–271 DOI 10.1016/j.biocon.2016.10.008.

da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, Maretty L, Zepeda-Mendoza ML, Campos PF, Heller R, Pereira RJ. 2016. Next-generation biology: sequencing and data analysis approaches for non-model organisms. *Marine Genomics* **30**:3–13 DOI 10.1016/j.margen.2016.04.012.

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34(15)**:2666–2669 DOI 10.1093/bioinformatics/bty149.

De Vivo M, Lee HH, Huang YS, Dreyer N, Fong CL, de Mattos FMG, Jain D, Wen YHV, Mwihaki JK, Wang TY, Machida RJ, Wang J, Chan BKK, Tsai IJ. 2022. Utilisation of Oxford Nanopore sequencing to generate six complete gastropod mitochondrial genomes as part of a biodiversity curriculum. *Scientific Reports* **12(1)**:118 DOI 10.1038/s41598-022-14121-0.

Díaz-Nieto JF, Jansa SA, Voss RS. 2016. Phylogenetic relationships of *Chacodelphys* (Marsupialia: Didelphidae: Didelphinae) based on "ancient" DNA sequences. *Journal of Mammalogy* **97(2)**:394–404 DOI 10.1093/JMAMMAL/GYV197.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32(5)**:1792–1797 DOI 10.1093/NAR/GKH340.

Elyasigorji Z, Izadpanah M, Hadi F, Zare M. 2023. Mitochondrial genes as strong molecular markers for species identification. *Nucleus (India)* **66(1)**:81–93 DOI 10.1007/S13237-022-00393-4.

Emser S, Schaschl H, Millesi E, Steinborn R. 2021. Extension of mitogenome enrichment based on single long-range PCR: mtDNAs and putative mitochondrial-derived peptides of five rodent hibernators. *Frontiers in Genetics* **12**:3616 DOI 10.3389/fgene.2021.685806.

Field R, Hawkins BA, Cornell HV, Currie DJ, Diniz-Filho JAF, Guégan JF, Kaufman DM, Kerr JT, Mittelbach GG, Oberdorff T, O'Brien EM, Turner JRG. 2009. Spatial species-richness gradients across scales: a meta-analysis. *Journal of Biogeography* **36(1)**:132–147 DOI 10.1111/j.1365-2699.2008.01963.x.

Filipović I, Rašić G, Hereward J, Gharuka M, Devine GJ, Furlong MJ, Etebari K. 2022. A high-quality de novo genome assembly based on nanopore sequencing of a wild-caught coconut rhinoceros beetle (*Oryctes rhinoceros*). *BMC Genomics* **23**:538 DOI 10.1186/S12864-022-08628-Z.

Franco-Sierra ND, Díaz-Nieto JF. 2020. Rapid mitochondrial genome sequencing based on Oxford Nanopore Sequencing and a proxy for vertebrate species identification. *Ecology and Evolution* **10(7)**:3544–3560 DOI 10.1002/ECE3.6151.

Galina CS, Martínez JF, Murphy BD. 2023. Constraints on research in biological and agricultural science in developing countries: the example of Latin America. *Publications* **11(2)**:22 DOI 10.3390/publications11020022.

Gan HM, Grandjean F, Jenkins TL, Austin CM. 2019. Absence of evidence is not evidence of absence: nanopore sequencing and complete assembly of the European lobster (*Homarus gammarus*) mitogenome uncovers the missing nad2 and a new major gene cluster duplication. *BMC Genomics* **20**:556 DOI 10.1186/s12864-019-5704-3.

Gaudioso A, Garcia-Rozas P, Casarejos MJ, Pastor O, Rodriguez-Navarro JA. 2019. Lipidomic alterations in the mitochondria of aged parkin null mice relevant to autophagy. *Frontiers in Neuroscience* **13**:3389 DOI 10.3389/FNINS.2019.00329.

Giarla TC, Voss RS. 2020. On the identity of Victorias Mouse Opossum Marmosa regina Thomas, 1898. *American Museum Novitates* **3960**:1–13 DOI 10.1206/3960.1.

Giménez AL, Giannini NP, Almeida FC. 2019. Mitochondrial genetic differentiation and phylogenetic relationships of three *Eptesicus* (*Histiotus*) species in a contact zone in Patagonia. *Mastozoologia Neotropical* **26(2)**:349–358 DOI 10.31687/saremMN.19.26.2.0.10.

Gostel MR, Kress WJ. 2022. The expanding role of DNA barcodes: indispensable tools for ecology, evolution, and conservation. *Diversity* **14(3)**:213 DOI 10.3390/D14030213.

Groenenberg DSJ, Harl J, Duijm E, Gittenberger E. 2017. The complete mitogenome of Orcula dolium (Draparnaud, 1801); Ultra-deep sequencing from a single long-range PCR using the Ion-Torrent PGM. *Hereditas* **154(1)**:1 DOI 10.1186/s41065-017-0028-2.

Gutiérrez EE, Anderson RP, Voss RS, Ochoa JG, Aguilera M, Jansa SA. 2014. Phylogeography of *Marmosa robinsoni*: insights into the biogeography of dry forests in northern South America. *Journal of Mammalogy* **95(6)**:1175–1188 DOI 10.1644/14-MAMM-A-069.

Gutiérrez EE, Jansa SA, Voss RS. 2010. Molecular systematics of mouse opossums (Didelphidae: *Marmosa*): assessing species limits using mitochondrial DNA sequences, with comments on phylogenetic relationships and biogeography. *American Museum Novitates* **3692**:1–22 DOI 10.1206/708.1.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270(1512)**:313–321 DOI 10.1098/rspb.2002.2218.

Hebert PDN, Ratnasingham S, de Waard JR. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences* **270(suppl_1)**:S96–S99 DOI 10.1098/rsbl.2003.0025.

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35(2)**:518–522 DOI 10.1093/MOLBEV/MSX281.

Imai K, Nemoto R, Kodana M, Tarumoto N, Sakai J, Kawamura T, Ikebuchi K, Mitsutake K, Murakami T, Maesaki S, Fujiwara T, Hayakawa S, Hoshino T, Seki M, Maeda T. 2020. Rapid and accurate species identification of mitis group streptococci using the MinION nanopore sequencer. *Frontiers in Cellular and Infection Microbiology* **10**:2–11 DOI 10.3389/FCIMB.2020.00011.

Jackman SD, Coombe L, Warren RL, Kirk H, Trinh E, MacLeod T, Pleasance S, Pandoh P, Zhao Y, Coope RJ, Bousquet J, Bohlmann J, Jones SJM, Birol I. 2020. Complete mitochondrial genome of a gymnosperm, Sitka Spruce (P*icea sitchensis*), indicates a complex physical structure. *Genome Biology and Evolution* **12(7)**:1174–1179 DOI 10.1093/GBE/EVAA108.

Johri S, Solanki J, Cantu VA, Fellows SR, Edwards RA, Moreno I, Vyas A, Dinsdale EA. 2019. 'Genome skimming' with the MinION hand-held sequencer identifies CITES-listed shark species in India's exports market. *Scientific Reports* **9**:e00590 DOI 10.1038/s41598-019-40940-9.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28(12)**:1647–1649 DOI 10.1093/BIOINFORMATICS/BTS199.

Kerfeld CA, Scott KM. 2011. Using BLAST to Teach "E-value-tionary" concepts. *PLOS Biology* **9(2)**:e1001014 DOI 10.1371/JOURNAL.PBIO.1001014.

Kneubehl AR, Muñoz-Leal S, Filatov S, de Klerk DG, Pienaar R, Lohmeyer KH, Bermúdez SE, Suriyamongkol T, Mali I, Kanduma E, Latif AA, Sarih M, Bouattour A, de León AAP, Teel PD, Labruna MB, Mans BJ, Lopez JE. 2022. Amplification and sequencing of entire tick mitochondrial genomes for a phylogenomic analysis. *Scientific Reports* **12**:19310 DOI 10.1038/s41598-022-23393-5.

Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nature Methods* **17(11)**:1103–1110 DOI 10.1038/s41592-020-00971-x.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37(5)**:540–546 DOI 10.1038/s41587-019-0072-8.

Kowalczyk M, Staniszewski A, Kamiñska K, Domaradzki P, Horecka B. 2021. Advantages, possibilities, and limitations of mitochondrial DNA analysis in molecular identification. *Folia Biologica (Poland)* **69**:101–111 DOI 10.3409/FB_69-3.12.

Krehenwinkel H, Pomerantz A, Prost S. 2019. Genetic biomonitoring and biodiversity assessment using portable sequencing technologies: current uses and future directions. *Genes* **10(11)**:858 DOI 10.3390/GENES10110858.

Kryukov AP, Spiridonova LN, Tyunin AP, Kryukov KA, Dorda BA. 2020. Complete mitochondrial genomes of five subspecies of the Eurasian magpie *Pica pica*, obtained with Oxford Nanopore MinION, and their interpretation regarding intraspecific taxonomy. *Mitochondrial DNA Part B: Resources* **5(3)**:3810–3811 DOI 10.1080/23802359.2020.1838354.

Kyle CJ, Wilson CC. 2007. Mitochondrial DNA identification of game and harvested freshwater fish species. *Forensic Science International* **166(1)**:68–76 DOI 10.1016/j.forsciint.2006.03.025.

Larsen RJ, Knapp MC, Genoways HH, Khan FAA, Larsen PA, Wilson DE, Baker RJ. 2012. Genetic diversity of neotropical *Myotis* (Chiroptera: Vespertilionidae) with an emphasis on South American species. *PLOS ONE* **7(10)**:e46578 DOI 10.1371/journal.pone.0046578.

Letunic I, Bork P. 2021. Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49(W1)**:W293–W296 DOI 10.1093/nar/gkab301.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34(18)**:3094–3100 DOI 10.1093/bioinformatics/bty191.

Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods* **12(8)**:733–735 DOI 10.1038/nmeth.3444.

Lopez-Baucells A, Rocha R, Tavares VD, Moras LM, Silva SE, Bobrowiec PED, Meyer CFJ. 2018. Molecular, morphological and acoustic identification of *Eumops maurus* and *Eumops hansae* (Chiroptera: Molossidae) with new reports from Central Amazonia. *Tropical Zoology* **31(1)**:1–20 DOI 10.1080/03946975.2017.1382284.

Maestri S, Cosentino E, Paterno M, Freitag H, Garces JM, Marcolungo L, Alfano M, Njunjic I, Schilthuizen M, Slik F, Menegon M, Rossato M, Delledonne M. 2019. A rapid and accurate MinION-based workflow for tracking species biodiversity in the field. *Genes* **10(6)**:468 DOI 10.3390/genes10060468.

Magi A, Giusti B, Tattini L. 2017. Characterization of MinION nanopore data for resequencing analyses. *Briefings in Bioinformatics* **18**:940–953 DOI 10.1093/BIB/BBW077.

Malukiewicz J, Cartwright RA, Dergam JA, Igayara CS, Nicola PA, Pereira LMC, Ruiz-Miranda CR, Stone AC, Silva DL, da Silva FDR, Varsani A, Walter L, Wilson MA, Zinner D, Roos C. 2021. Genomic skimming and nanopore sequencing uncover cryptic hybridization in one of world's most threatened primates. *Scientific Reports* **11**:525 DOI 10.1038/s41598-021-96404-6.

Meirmans PG. 2012. The trouble with isolation by distance. *Molecular Ecology* **21(12)**:2839–2846 DOI 10.1111/J.1365-294X.2012.05578.X.

Meng G, Li Y, Yang C, Liu S. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research* **47(11)**:e63 DOI 10.1093/nar/gkz173.

Milan DT, Mendes IS, Damasceno JS, Teixeira DF, Sales NG, Carvalho DC. 2020. New 12S metabarcoding primers for enhanced Neotropical freshwater fish biodiversity assessment. *Scientific Reports* **10**:17966 DOI 10.1038/s41598-020-74902-3.

Moore WS. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* **49**:718–726 DOI 10.1111/j.1558-5646.1995.tb02308.x.

Moratelli R, Gardner AL, De Oliveira JA, Wilson DE. 2013. Review of *Myotis* (Chiroptera, Vespertilionidae) from northern South America, including description of a new species. *American Museum Novitates* **3780**:1–36 DOI 10.1206/3780.2.

Moratelli R, Wilson D, Novaes RL, Helgen K, Gutiérrez E. 2017. Caribbean *Myotis* (Chiroptera, Vespertilionidae), with description of a new species from Trinidad and Tobago. *Journal of Mammalogy* **98(4)**:1–15 DOI 10.1093/jmammal/gyx062.

Moreno CE, Calderón-Patrón JM, Martín-Regalado N, Martínez-Falcón AP, Ortega-Martínez IJ, Rios-Díaz CL, Rosas F. 2018. Measuring species diversity in the tropics: a review of methodological approaches and framework for future studies. *Biotropica* **50(6)**:929–941 DOI 10.1111/btp.12607.

Nachman MW, Beckman EJ, Bowie RC, Cicero C, Conroy CJ, Dudley R, Hayes TB, Koo MS, Lacey EA, Martin CH, McGuire JA, Patton JL, Spencer CL, Tarvin RD, Wake MH, Wang IJ, Achmadi A, Álvarez-Castañeda ST, Andersen MJ, Arroyave J, Austin CC, Barker FK, Barrow LN, Barrowclough GF, Bates J, Bauer AM, Bell KC, Bell RC, Bronson AW, Brown RM, Burbrink FT, Burns KJ, Cadena CD, Cannatella DC, Castoe TA, Chakrabarty P, Colella JP, Cook JA, Cracraft JL, Davis DR, Davis Rabosky AR, D'Elía G, Dumbacher JP, Dunnum JL, Edwards SV, Esselstyn JA, Faivovich J, Fjeldså J, Flores-Villela OA, Ford K, Fuchs Jérôme, Fujita MK, Good JM, Greenbaum E, Greene HW, Hackett S, Hamidy A, Hanken J, Haryoko T, Hawkins MTR, Heaney LR, Hillis DM, Hollingsworth BD, Hornsby AD, Hosner PA, Irham M, Jansa S, Jiménez RA, Joseph L, Kirchman JJ, LaDuc TJ, Leaché AD, Lessa EP, López-Fernández H, Mason NA, McCormack JE, McMahan CD, Moyle RG, Ojeda RA, Olson LE, Kin Onn C, Parenti LR, Parra-Olea G, Patterson BD, Pauly GB, Pavan SE, Peterson AT, Poe S, Rabosky DL, Raxworthy CJ, Reddy S, Rico-Guevara A, Riyanto A, Rocha LA, Ron SR, Rovito SM, Rowe KC, Rowley J, Ruane S, Salazar-Valenzuela D, Shultz AJ, Sidlauskas B, Sikes DS, Simmons NB, Stiassny MLJ, Streicher JW, Stuart BL, Summers AP, Tavera J, Teta P, Thompson CW, Timm RM, Torres-Carvajal O,

Voelker G, Voss RS, Winker K, Witt C, Wommack EA, Zink RM. 2023. Specimen collection is essential for modern science. *PLOS Biology* **21(11)**:e3002318 DOI 10.1371/journal.pbio.3002318.

Naidu A, Fitak RR, Munguia-Vega A, Culver M. 2012. Novel primers for complete mitochondrial cytochrome b gene sequencing in mammals. *Molecular Ecology Resources* **12(2)**:191–196 DOI 10.1111/j.1755-0998.2011.03078.x.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32(1)**:268–274 DOI 10.1093/molbev/msu300.

Ni Y, Liu X, Simeneh ZM, Yang M, Li R. 2023. Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. *Computational and Structural Biotechnology Journal* **21(20)**:2352–2364 DOI 10.1016/j.csbj.2023.03.038.

Nowak C, Buntjen M, Steyer K, Frosch C. 2014. Testing mitochondrial markers for noninvasive genetic species identification in European mammals. *Conservation Genetics Resources* **6(1)**:41–44 DOI 10.1007/s12686-013-0068-5.

Page TJ, Hughes JM. 2010. Comparing the performance of multiple mitochondrial genes in the analysis of Australian freshwater fishes. *Journal of Fish Biology* **77(9)**:2093–2122 DOI 10.1111/J.1095-8649.2010.02821.X.

Parker J, Helmstetter AJ, Devey D, Wilkinson T, Papadopulos AST. 2017. Field-based species identification of closely-related plants using real-time nanopore sequencing. *Scientific Reports* **7(1 7)**:1–8 DOI 10.1038/s41598-017-08461-5.

Parson W, Pegoraro K, Niederstätter H, Föger M, Steinlechner M. 2000. Species identification by means of the cytochrome b gene. *International Journal of Legal Medicine* **114**:23–28 DOI 10.1007/S004140000134/METRICS.

Parsons DJ, Pelletier TA, Wieringa JG, Duckett DJ, Carstens BC. 2022. Analysis of biodiversity data suggests that mammal species are hidden in predictable places. *Proceedings of the National Academy of Sciences of the United States of America* **119(14)**:e2103400119 DOI 10.1073/pnas.2103400119.

Patton JL, Pardiñas UFJ, D'Elía G. 2015. *Mammals of South America.* Vol. 2. Chicago: University of Chicago Press DOI 10.7208/chicago/9780226169606.001.0001.

Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. 2018. Long reads: their purpose and place. *Human Molecular Genetics* **27(R2)**:R234–R241 DOI 10.1093/hmg/ddy177.

Pomerantz A, Peñafiel N, Arteaga A, Bustamante L, Pichardo F, Coloma LA, Barrio-Amorós CL, Salazar-Valenzuela D, Prost S. 2018. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *GigaScience* **7(4)**:1–14 DOI 10.1093/gigascience/giy033.

Pomerantz A, Sahlin K, Vasiljevic N, Seah A, Lim M, Humble E, Kennedy S, Krehenwinkel H, Winter S, Ogden R, Prost S. 2022. Rapid in situ identification of biological specimens via DNA amplicon sequencing using miniaturized laboratory equipment. *Nature Protocols* **17(6)**:1415–1443 DOI 10.1038/s41596-022-00682-x.

Ramón-Laca A, Gallego R, Nichols KM. 2023. Affordable de novo generation of fish mitogenomes using amplification-free enrichment of mitochondrial DNA and deep sequencing of long fragments. *Molecular Ecology Resources* **23(4)**:818–832 DOI 10.1111/1755-0998.13758.

Rastogi G, Dharne MS, Walujkar S, Kumar A, Patole MS, Shouche YS. 2007. Species identification and authentication of tissues of animal origin using mitochondrial and nuclear markers. *Meat Science* **76(4)**:666–674 DOI 10.1016/J.MEATSCI.2007.02.006.

Riaz N, Leung P, Barton K, Smith MA, Carswell S, Bull R, Lloyd AR, Rodrigo C. 2021. Adaptation of Oxford Nanopore technology for hepatitis C whole genome sequencing and identification of within-host viral variants. *BMC Genomics* **22**:148 DOI 10.1186/s12864-021-07460-1.

Rios CA, de Regules S. 2022. Responsible science communication in Latin America: reflections on challenges. *Journal of Science Communication* **21(04)**:C03 DOI 10.22323/2.21040303.

Rodríguez-Posada ME, Morales-Martínez DM, Ramírez-Chaves HE, Martínez-Medina D, Calderón-Acevedo CA. 2021. A new species of Long-eared Brown Bat of the genus *Histiotus* (Chiroptera) and the revalidation of *Histiotus colombiae*. *Caldasia* **43(2)**:221–234 DOI 10.15446/caldasia.v43n2.85424.

Rossi RV, Voss RS, Lunde DP. 2010. A revision of the Didelphid marsupial genus *Marmosa* part 1. The species in tate's "mexicana" and "mitis" sections and other closely related forms. *Bulletin of The American Museum of Natural History* **334**:1–83 DOI 10.1206/334.1.

Ruelas D, Pacheco V. 2022. Variación intraespecífica de *Carollia brevicaudum* y *Carollia perspicillata* (Phyllostomidae: Chiroptera) de Perú y Ecuador. *Neotropical Biodiversity* **8(1)**:253–266 DOI 10.1080/23766808.2022.2093816.

Ruelas D, Pacheco V, Inche B, Tinoco N. 2021. A preliminary review of Nephelomys albigularis (Tomes, 1860) (Rodentia: Cricetidae), with the description of a new species from the Peruvian montane forests. *Zootaxa* **5027(2)**:175–210 DOI 10.11646/ZOOTAXA.5027.2.3.

Sa W, Shang QH, Yang T, Gao QY, Liu ML, Liang J, Li ZH. 2020. Characterization of the complete mitochondrial genome of *Morchella eohespera*. *Mitochondrial DNA Part B* **5(3)**:3066–3067 DOI 10.1080/23802359.2020.1797559.

Salager-Meyer F. 2008. Scientific publishing in developing countries: challenges for the future. *Journal of English for Academic Purposes* **7(2)**:121–132 DOI 10.1016/J.JEAP.2008.03.009.

Sanderson ND, Kapel N, Rodger G, Webster H, Lipworth S, Street TL, Peto T, Crook D, Stoesser N. 2023. Comparison of R9.4.1/Kit10 and R10/Kit12 Oxford Nanopore flowcells and chemistries in bacterial genome reconstruction. *Microbial Genomics* **9(1)**:mgen000910 DOI 10.1099/MGEN.0.000910.

Schlegel M, Ali HS, Stieger N, Groschup MH, Wolf R, Ulrich RG. 2012. Molecular identification of small mammal species using novel Cytochrome b gene-derived degenerated primers. *Biochemical Genetics* **50(5–6)**:440–447 DOI 10.1007/s10528-011-9487-8.

Schäffer S, Zachos FE, Koblmüller S. 2017. Opening the treasure chest: a DNA-barcoding primer set for most higher taxa of Central European birds and mammals from museum collections. *PLOS ONE* **12(3)**:e0174449 DOI 10.1371/journal.pone.0174449.

Shah N, Nute MG, Warnow T, Pop M. 2019. Misunderstood parameter of NCBI BLAST impacts the correctness of bioinformatics workflows. *Bioinformatics* **35(9)**:1613–1614 DOI 10.1093/BIOINFORMATICS/BTY833.

Sikes RS, Bryan JA. 2016. Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *Journal of Mammalogy* **97(3)**:663–688 DOI 10.1093/jmammal/gyw078.

Srivathsan A, Hartop E, Puniamoorthy J, Lee WT, Kutty SN, Kurina O, Meier R. 2019. Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology* **17(1)**:e1001130 DOI 10.1186/s12915-019-0706-9.

Srivathsan A, Lee L, Katoh K, Hartop E, Kutty SN, Wong J, Yeo D, Meier R. 2021. ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. *BMC Biology* **19**:313 DOI 10.1186/s12915-021-01141-x.

Stockburger C, Miano D, Baeumlisberger M, Pallas T, Arrey TN, Karas M, Friedland K, Muller WE. 2016. A mitochondrial role of SV2a protein in aging and alzheimer's disease: studies with levetiracetam. *Journal of Alzheimer's Disease: JAD* **50(1)**:201–215 DOI 10.3233/JAD-150687.

Stoekle M. 2003. Taxonomy, DNA, and the bar code of life. *BioScience* **53**:796–797 DOI 10.1641/0006-3568(2003)053.

Teta P, Díaz-Nieto JF. 2019. How integrative taxonomy can save a species from extinction: the supposedly extinct mouse opossum *Cryptonanus ignitus* (Diaz, Flores and Barquez, 2000) is a synonym of the living C. chacoensis (Tate, 1931). *Mammalian Biology* **96**:73–80 DOI 10.1016/j.mambio.2019.04.004.

Velazco PM, Patterson BD. 2013. Diversification of the Yellow-shouldered bats, Genus *Sturnira* (Chiroptera, Phyllostomidae), in the New World tropics. *Molecular Phylogenetics and Evolution* **68(3)**:683–698 DOI 10.1016/j.ympev.2013.04.016.

Voss RS, Emmons LH. 1996. Mammalian diversity in neotropical lowland rainforests: a preliminary assessment. *Bulletin of the American Museum of Natural History* **230**:1–115.

Voss RS, Giarla TC, Diáz-Nieto JF, Jansa SA. 2020. A revision of the Didelphid Marsupial Genus *Marmosa* Part 2. Species of the *Rapposa* Group (Subgenus *Micoureus*). *Bulletin of the American Museum of Natural History* **439(1)**:3–60 DOI 10.1206/0003-0090.439.1.1.

Weksler M, Percequillo AR, Voss RS. 2006. Ten new genera of Oryzomyine rodents (Cricetidae: Sigmodontinae). *American Museum Novitates* **3537**:1–29 DOI 10.1206/0003-0082(2006)3537.

Wiley G, Miller MJ. 2020. A highly contiguous genome for the golden-fronted woodpecker (*Melanerpes aurifrons*) via hybrid oxford nanopore and short read assembly. *G3: Genes, Genomes, Genetics* **10(6)**:1829–1836 DOI 10.1534/g3.120.401059.

Wright S. 1943. Isolation by distance. *Genetics* **28(2)**:114–138 DOI 10.1093/GENETICS/28.2.114.

Yim WC, Cushman JC. 2017. Divide and conquer (DC) BLAST: fast and easy BLAST execution within HPC environments. *PeerJ* **2017**:e3486 DOI 10.7717/PEERJ.3486/TABLE-1.

Zarog M. 2023. Improving economic research impact in low-income countries: key challenges and the way forward. *Indian Journal of Economics and Development* **19**:216–223 DOI 10.35716/IJED/22219.

Zascavage RR, Hall CL, Thorson K, Mahmoud M, Sedlazeck FJ, Planz JV. 2019. Approaches to whole mitochondrial genome sequencing on the oxford nanopore minion. *Current Protocols in Human Genetics* **104(1)**:e94 DOI 10.1002/cphg.94.

Zhang W, Cui H, Wong LJC. 2012. Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clinical Chemistry* **58(9)**:1322–1331 DOI 10.1373/clinchem.2011.181438.

Zhang X, Liu CG, Yang SH, Wang X, Bai FW, Wang Z. 2022. Benchmarking of long-read sequencing, assemblers and polishers for yeast genome. *Briefings in Bioinformatics* **23(3)**:345 DOI 10.1093/BIB/BBAC146.