International Journal of Primatology
https://doi.org/10.1007/s10764-024-00429-3

Overcoming Challenges to Extracting and Sequencing Historical DNA to Support Primate Evolutionary Research and Conservation, with an Application to Galagos

Anna Penna[1,2] · Mary E. Blair[3] · Hsiao-Lei Lui[2] · Elsa Peters[2,4] · Logan Kistler[2] · Luca Pozzi[1]

Received: 21 April 2023 / Accepted: 14 March 2024

Abstract

The availability of genetic data from wild populations limits our understanding of primate evolution and conservation, particularly for small nocturnal species such as lorisiforms (galagos, lorises, angwantibos, and pottos). Emerging methods for recovering genomic DNA from historical museum specimens have been rarely used in primate studies. We aimed to optimize extraction and bioinformatics protocols to maximize the recovery of historical DNA to fill important geographic and taxonomic gaps, improve phylogenetic resolution, and inform conservation of Lorisiform primates. First, we compared the performance of two DNA extraction methods by using 238 specimens up to a hundred years old. We then selected 96 samples with the highest DNA yields for shotgun sequencing. To evaluate the impact of phylogenetic divergence in bioinformatic read mapping, we compared coverage depths when using human and three lorisiform reference mitogenomes. Based on whole genomic data, we performed metagenomics and microbial diversity analyses to assess the composition of potentially exogenous content. Lastly, based on the most geographically and taxonomically comprehensive sampling for the West African lorisiforms to date (19/32 currently recognized species), we performed phylogenetic inference using Maximum Likelihood. The results showed that older samples yield lower DNA concentration, with an optimized phenol-chloroform protocol outperforming a commercial kit. However, both extraction methods generated DNA in sufficient amount and quality for phylogenetic inference. Our reference bias comparisons showed that higher phylogenetic proximity between focal species and reference mitogenome increases coverage depth. The metagenomic analysis found human contamination in only one of 96 samples (1%), whereas ten of 96 (11%) samples showed nonnegligible levels of other exogenous contents, among which are certain blood parasites. We inferred low support for the monophyly of Asian and African Lorisids but confirmed

Handling Editor: Frank Cuozzo

Extended author information available on the last page of the article

Published online: 21 May 2024

the monophyly and previously suggested relationships among Galagid genera. Lastly, we found evidence of cryptic species diversity within the western dwarf galagos (genus *Galagoides*). Taken together, these results attest to the enormous potential of museomics to advance our understanding of galago evolution, ecology, and conservation, an approach that can be extended to other primate clades.

Keywords Museum specimens · Genetics · Museomics · Metagenomics · Parasites · Reference bias · *Galagoides*

Introduction

Primate molecular phylogenetics has always been in the front of the development of novel molecular techniques (from serological data and protein electrophoresis to genomics and proteomics; Yoder 2013; Ting & Sterner 2013). Genetic data have revolutionized our understanding of the biogeographical history and evolutionary relationships within primates (Fabre *et al*, 2009; Perelman *et al*., 2011; Springer et al., 2012; Pozzi *et al*., 2014a; Roger & Gibbs, 2014), including the colonization of Madagascar by the ancestor of Lemuriformes and *Daubentonia madagascariensis*; Yoder, 1996), the phylogenetic positioning of *Homo sapiens* and African apes in the same family to the exclusion of the genus *Pongo*; Goodman et al., 1998), and the closer relationship of Tarsiiformes to Siimiformes than to tooth-combed primates (Strepsirrhini; Jameson et al., 2011). Although some clades were the focus of much molecular phylogenetic scrutiny, others have historically remained in the periphery of genetic studies. This is especially true for nocturnal taxa, such as the Lorisiformes (an infraorder of Strepsirrhini that includes galagos, lorises, pottos, and angwantibos). When coupled with geographic, morphological, behavioral, and fossil data, recent genetic studies have allowed the resolution of deep-divergence estimates (Munds *et al*., 2018; Pozzi *et al*., 2014b; 2015), description of new genera (*Paragalago*: Masters *et al*., 2017; *Xanthonycticebus*: Nekaris & Nijman, 2022), documentation of cryptic diversity (Blair *et al*., 2023; Penna *et al*., 2022), and reconstruction of the biogeographic history of Lorisiformes (Pozzi, 2016). Traditionally, most primate genetic studies have relied on high-quality fresh tissue (e.g., muscle, liver, ear clip, and blood) from living animals collected in the field or captivity. However, to obtain high-quality tissue samples, researchers must capture, manipulate, and sometimes sedate the animals. This task is not only extremely challenging when sampling free-ranging populations, but often requires a trained and licensed professional, with regulations varying dramatically between countries (Glander, 2013).

To overcome financial and logistic challenges and minimize the danger and behavioral impact of trapping or darting live animals in the wild, primatologists can sample genetic material from free-ranging animals using alternative noninvasive strategies (Waits & Paetkau, 2005). The most used noninvasive sources of DNA are fecal (Ang *et al*., 2020; Frantzen *et al*., 1998; Parker *et al*., 2022), hair (Woodruff, 1993), and urine samples (Ozga *et al*., 2021), or food traces (Aylward *et al*., 2018). Promising bioinformatic methodological improvements also allow for more

precise quantification of genotyping error rates associated with these low-quality sampling strategies (Bonin *et al*., 2004). Genetic data also can be obtained from biological material sampled in association with local communities, for instance, from animals brought back by hunters (Ruiz-García *et al*. 2018) or salvaged from illegal trade (Brown *et al*. 2014). However, relying on tissue collected from these carcasses comes with the caveat of rapid post-mortem DNA degradation and a bias toward species that are more likely to be hunted or tend to fall more often into traps. In nocturnal primates, it is virtually impossible to rely on fecal or urine samples, and most species are rarely found in markets (Wilkie & Carpenter, 1999; Fa *et al*., 2006; Fominka *et al*., 2021). Moreover, trapping smaller-sized nocturnal species is particularly challenging in the wild, because they tend to be averse to the trapping apparatus (e.g., Sherman or Tomahawk traps) and are rarely attracted by regular baits (e.g., fruits, peanut butter, and fish oil paste). Consequently, very few studies have collected fresh tissue samples in the field, usually sampling within restricted geographic ranges (Penna *et al*., 2022; Phukuntsi *et al*., 2019, 2021; Pozzi, Penna *et al*., 2020a, Pozzi *et al*., 2019).

To circumvent these limitations, researchers have turned to historical specimens housed in natural history museums as sources of genetic samples (Rowe *et al*., 2011; Burrell *et al*., 2015; McCormack *et al*., 2016; Gonçalves *et al*., 2021; Parker *et al*., 2020). Because of their completeness and preservation, these institutions have played a central role in anatomy and taxonomic classification studies over the last three centuries (Mauriès, 2002). In extreme cases, natural history collections have become the only remaining source of DNA data for many rare, endangered, and extinct species, providing a crucial source of information for comparative and temporal analyses (Diez-del-Molino *et al*., 2018; Schmitt *et al*., 2019; Castañeda-Rico *et al*., 2020; Roycroft *et al*., 2022). Historical specimens can be as old as 200 years and usually consist of body parts or entirely preserved nonliving animals, coupled with ancillary information (e.g., sex, age, locality, the context of the collection, and body measurements while alive). Because preservation methods often rely on chemical treatments (e.g., potassium alum, sodium chloride, and sulfuric acid; Bock & Quaiser, 2019) and the cells in the dead bodies of these preserved specimens lack self-regulatory mechanisms (DNA repair machinery), postmortem DNA degradation accumulates over time, leading to changes in DNA structure (Raxworthy & Smith, 2021). Fortunately, recent advances in laboratory protocols and sequencing methods have allowed for the characterization of highly fragmented DNA molecules, typical of historical samples housed in natural history collections.

These methodological advances have unlocked an unforeseen genetic potential of biological specimens collected decades or centuries ago, thereby increasing the value of natural history collections for biological inquiry. The field of museomics (Card *et al*., 2021; Raxworthy & Smith, 2021), emerged as a subfield of ancient DNA (Shapiro & Hofreiter, 2014) that seeks to retrieve genetic data from museum specimens. Despite these promises, the process of obtaining genetic material from historical specimens remains challenging and prone to failure, mostly due to sample age and previous chemical treatment during the preparation process. Consequently, the remaining DNA is typically low in quantity and highly fragmented. The recent development of massively parallel short-read sequencing in the past two decades,

referred to as next-generation sequencing (Burrell *et al*., 2015; Der Sarkissian *et al*., 2015), have overcome some of these challenges in the past two decades. Although providing opportunities to access novel data types from hard-to-access species, these methods have also raised concerns. First, some destructive sampling is necessary, which is at some level of disagreement with museums' long-term goals of specimen care (Freedman *et al*., 2018). Second, historical samples are prone to DNA contamination from sources different than the target specimen, such as bacteria, fungi, insects, and even humans. To circumvent these challenges, scientists have developed sampling, laboratory, and bioinformatics protocols that maximize DNA yields and quality while ensuring minimal damage to voucher specimen structures and minimal contamination from exogenous DNA sources. Researchers have attempted different solutions to accommodate the diversity of specimen preparation types with variable success (McDonough *et al*., 2018), and museomics has become an active research area in and of itself (Buerki & Baker, 2016; Card *et al*., 2021; Hawkins, Bailey *et al*., 2022a, Hawkins, Flores *et al*., 2022b; Raxworthy & Smith, 2021).

DNA retrieved from museum specimens has supported studies of species identification and geographic distribution, complementing traditional approaches in taxonomy (Blair *et al*., 2023; Hawkins, Bailey *et al*., 2022a, Hawkins, Flores *et al*., 2022b). Researchers have used historical DNA data to infer the evolutionary relationships of broadly distributed, rare, or inaccessible species, clarifying the taxonomic status of these species and allowing for the resolution of taxonomic problems, taxa distribution range, description of new species, and the elucidation of evolutionary processes underlying speciation (Ennes Silva *et al*., 2022; Guschanski *et al*., 2013; Hawkins, Bailey *et al*., 2022a; Porter *et al*., 2021; Yao *et al*., 2020; Jensen *et al*., 2023). The combination of historical and contemporary samples also has allowed for the characterization of changes in the genetic composition of natural populations over time (Clark *et al*., 2023). Researchers have used this approach to identify species responses to environmental shifts and declining populations of conservation concern (van der Valk *et al*., 2019) and study recent extinctions (Woods *et al*., 2018; Kistler *et al*., 2015; Marciniak *et al*., 2021). Natural history collections provide novel opportunities for hypothesis-driven research beyond phylogenetic resolution, mimicking experimental designs traditionally employed in laboratory settings. In doing so, they reaffirm the unique position of natural history collections in identifying issues in biological conservation and supporting the development of solutions to these issues.

Leveraging DNA from historical specimens may offer increased potential to enhance our understanding of evolutionary relationships and species diversity among the Lorisiforms, which remain some of the least studied primates (Nekaris & Burrows, 2020; Pozzi *et al*., 2015). Currently, there are two subfamilies recognized within the Lorisiformes: Lorisidae, which includes the subfamilies Lorisinae (Asian lorises) and Perodicticinae (African lorises), and Galagidae, which are the galagos located in Sub-Saharan Africa. The Galagidae family is monophyletic (Perelman *et al*., 2011; Pozzi *et al*., 2014b, 2015; Springer *et al*., 2012), with its divergence dating back to approximately 30 Ma at the beginning of the Oligocene in Central Africa (Pozzi *et al*., 2014b, 2015; Pozzi, 2016). Some studies have supported a sister relationship between the two Lorisidae genera found in Africa (*Arctocebus, Perodicticus*) and the three found in Asia (*Loris, Nycticebus, Xanthonycticebus*) (Pozzi *et al*., 2015; Nekaris & Nijman,

2022), which suggests that two distinct lorisiform radiations occurred in the African continent. However, many studies have not found support for the monophyly of Lorisidae (Springer *et al*., 2012; Pozzi *et al*., 2014b), likely because of poor taxon representation across groups (due to the many difficulties of sampling nocturnal primates, as discussed above), and because it is likely that rapid diversification occurred early in the lineage (Springer *et al*., 2012; Pozzi *et al*., 2014b, Pozzi, Penna *et al*., 2020a, Pozzi, Roos *et al*., 2020b). Recent studies of species diversity within Lorisiforms suggest that our account of species richness for many genera in this group remains underestimated (Pozzi *et al*., 2020a; Penna *et al*., 2022; Blair *et al*., 2023).

In this contribution, we explored the potential of incorporating historical DNA from lorisiform museum specimens to improve the genetic analyses of these understudied and hard-to-sample nocturnal primates. We hope that our study inspires more primatologists to incorporate genetic information from museum specimens in their studies. Our approach consisted of five steps. First, we sampled tissue from 238 specimens deposited in four natural history collections. These specimens represented 19 nominal species-level taxa from two families: Lorisidae and Galagidae. Second, we employed two extraction protocols with different degrees of complexity and customization to assess the molecular laboratory procedures that maximize DNA yield and to understand the effects of specimen age on extraction success. Third, we explored the potential of metagenomics analyses on museum specimens in detecting blood-borne parasites and other postmortem exogenous contaminants that are likely associated with long-term preservation in museum cabinets. Fourth, we examined the impact of reference bias, that is the evolutionary divergence between the focal species and the species serving as a reference on mapping success, to determine which reference maximizes the number of reads mapped and genomic region coverage. Lastly, we tested the potential of archival DNA to support phylogenetic inference at both shallow and deep levels of evolutionary divergence by characterizing patterns of genetic variation within and between understudied lorisiform taxa, with a particular emphasis on western dwarf galagos (*Galagoides*).

Methods

Tissue Sampling

We sampled 238 specimens housed in four natural history collections: the American Museum of Natural History, New York (AMNH, N = 64); the Smithsonian Institution National Museum of Natural History, Washington, D.C. (USNM, N = 63); the Field Museum of Natural History (Chicago, IL, N = 20); and the Royal Museum of Central Africa, Belgium (RMCA, N = 91). Tissue sampling followed a sterile protocol designed to minimize damage to museum specimens and the chance of external contamination. We used disposable feather scalpels and blades to obtain thin slices of tissue from dry preparations ("study skins"). We covered the working surfaces with disposable wipe sheets, replacing them after each tissue collection. Using gloves and sterile tweezers, we transferred the sampled tissue to labeled screw-cap tubes for storage in dry conditions until DNA extraction.

To minimize the risk of exogenous contamination during DNA extraction and library preparation, we then transferred all samples to a historic DNA facility where no polymerase chain reactions (PCR) are allowed (Fig. 1).

We targeted body parts that are not associated with any taxonomic diagnostic traits, specifically the lateral edge of the left hindlimb big toe pad. Compared with other tissues usually found in museum collections (e.g., bone, skin, and hair), several studies have shown that toepads yield higher PCR amplification rates, lower genotyping errors, and minimal specimen destruction (Lonsinger *et al*., 2019; Polanc *et al*., 2012). In the few instances when skins were not available, we collected excess adherent tissue from skeletal parts ("osteocrusts," such as brain or muscle). A detailed list of the specimen information is available in the Supporting Information S1.

DNA Extraction and Purification

We compared two DNA extraction protocols with different complexities. Protocol 1 consisted of a phenol-chloroform extraction protocol modified to maximize the final concentration of endogenous DNA, retain shorter fragments, and minimize contamination from exogenous DNA content, following Campos & Gilbert (2019) and McDonough *et al*. (2018) (Supporting Information File S1). In summary, we

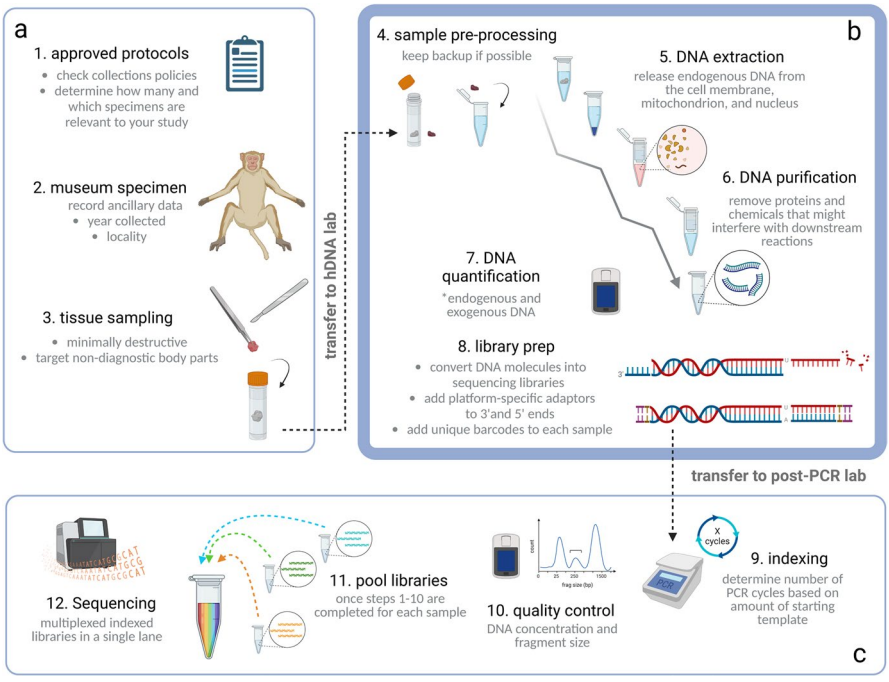

Fig. 1 Museomics workflow in a study of historical primate specimens. The main steps include sample collection, DNA extraction, library preparation, and Illumina sequencing. Boxes show activities in (**a**) museum collections, (**b**) historical or ancient molecular biology laboratories, and (**c**) post-PCR laboratory. Figure created with BioRender.com, license provided in Supplementary File 1.

washed each individual sample in 150 µL of 0.05% bleach solution for 1 min followed by two 1-min washes with double-distilled water to remove bleach residues. Bleach pretreatment is an effective method for reducing microbial and human contamination, thus increasing the endogenous DNA content (Korlević *et al*., 2015). We put the washed samples in a lysing buffer containing sodium dodecyl sulfate (SDS), dithiothreitol (DTT), and Proteinase K at 56 °C in a shaking incubator for overnight digestion. We then performed a round of 1:1 phenol-chloroform organic extraction, followed by MinElute silica membrane column spinning (Qiagen cat. No. 28004) with 5X binding buffer for DNA purification and final elution in Qiagen EB buffer after incubation at 37 °C for 20 min. We used Protocol 1 to extract samples from AMNH, FMNH, and USNM collections.

Protocol 2 is a modification of a commercially available kit commonly used to extract high-quality DNA from modern samples: the Qiagen DNEasy Blood and Tissue Kit (Qiagen cat. No. 69504). This protocol is shorter than Protocol 1 and does not have the bleach washing step before lysis. Briefly, we used a solution of Proteinase K, DTT, and Qiagen ATL lysis buffer to digest samples overnight in at 56 °C in a shaking incubator. After digestion, we followed the manufacturer's instructions, including a spin column purification and the final elution in a heated Qiagen AE buffer solution. We used Protocol 2 to extract DNA only from samples from RMCA.

Both protocols include a purification step using silica membrane column spin to ensure the removal of any potential reaction inhibitors in downstream DNA libraries. However, the minimum fragment sizes retained by these columns differed for each protocol. The MinElute column used in Protocol 1 could retain fragments as small as 70 bp, whereas the columns provided with the DNEasy kit used in Protocol 2 could not retain fragments smaller than 100 bp (according to the manufacturer). We performed all extractions in small batches (4–16 samples) including one negative control after every eight samples. We quantified the final DNA concentrations of all eluted extracts and controls using a Qubit fluorometer (1x ds DNA High-Sensitivity assay). Detailed protocols are available in the Supporting Information (Supplemental File 2).

We calculated the final amount of total DNA (ng) as the product between DNA concentration and elution volumes of the samples. To test whether sample age predicts the amount of final DNA, we evaluated the strength of the association between sample age and the amount of DNA obtained when using the different protocols using Pearson's correlation tests. To test for a difference in the performance of the two protocols, we compared the mean DNA yield per sample, mean fragment length, and redundancy level between the two protocols using parametric *t*-tests with a significance level of 0.05. We used R (R Core Team, 2022) for all statistical analysis and data visualization using the package "tidyverse" (Wickham *et al*. 2019)**.**

A dedicated ancient/historical DNA laboratory has become the standard for DNA extraction from museum specimens (Fulton & Shapiro, 2019). We performed all laboratory experiments (sample preparation, DNA extraction, and library preparation) inside the PCR hood at controlled facilities at the Smithsonian National Museum of Natural History Museum Support Center, Maryland, USA (Protocol 1), and the American Museum of Natural History, New York (Protocol 2). Both facilities

enforce strict use of personal protective equipment (PPE, e.g., full-body suit, double glove, mask, hair net) and controlled ventilation system (e.g., filters and negative air pressure) designed to impose unidirectional airflow that helps minimize exogenous contamination.

Library Preparation and Sequencing

We selected 96 samples with high DNA concentrations from the USNM and RMCA collections for whole-genome shotgun sequencing. We prepared individual libraries following the double-strand Blunt End Single Tube (BEST) protocol, which was designed to optimize the preparation of degraded samples for sequencing on Illumina platforms (Carøe *et al*., 2017). Compared with other library preparation protocols, BEST has two main advantages for dealing with degraded samples. First, it eliminates intermediate cleaning steps that often lead to the loss of shorter fragments (<70 bp), which are more common in the initial templates of degraded samples. Second, it uses enzymes developed for heat inactivation at 65 °C, as small fragments (<25 bp) are known to denature at 72 °C, depending on the sequence composition (Carøe *et al*., 2017).

Adapter dimers formed during library preparation can dominate the sequencing run, and consequently, decrease library sequencing yields. To minimize the formation of unwanted adapter dimers and prevent enzymatic activity from occurring before setting up the reactions, we completed the library preparations on a cooling block, adding enzymes at the end of each step. Library preparation involved the following steps: First, we diluted individual DNA templates to 16 or 32 μL starting volumes, and treated libraries with New England Biosystems (NEB) enzymes for three sequential reactions: 1) end-repair to digest 3' blunt-ends; 2) ligation of complementary strands of the 5' overhangs, inserting adenosine to the opposite strands containing uracil bases from DNA degradation; and 3) filling the space between the complementary strands. Finally, we added unique combinations of identification barcodes to individual during indexing, and annealed P5 and P7 primers to each sample for subsequent multiplexing. Following the recommendations of Meyer & Kircher (2010), we performed adapter indexing through 12 or 16 PCR cycles, depending on the starting template concentration (which ranged from 10 to 200 ng). To remove potential adapter dimers, we cleaned each individual library using 1.5 × volume of SPRI magnetic beads before and after the indexing PCR step. A detailed protocol for library preparation and indexing is available in Supplementary File 3. We prepared all libraries at the National Museum of Natural History Historical DNA Laboratory located at the Smithsonian Museum Support Center, Maryland.

Following library preparation, we used a Qubit fluorometer (1x ds High Sensitivity) to estimate the final DNA concentration (ng/μL) and visualized the DNA fragment distribution (base pairs, bp) in an Agilent TapeStation (High Sensitivity D1000). We used the following formula to calculate the final library concentration (nM): (concentration (ng/μL)/(660 (g/mL) × mean size (bp))) × 10^6. We then pooled individual libraries at equimolar concentrations prior to multiplexed sequencing in

two separate lanes of an Illumina HiSeq 2 × 150 system, containing 48 pooled samples in each lane. We shipped the pooled samples for sequencing at Admera Health (South Plainfield, NJ).

Raw Reads Processing

We removed sequencing adapters and unique barcodes from the 3' and 5' ends using Adapter Removal v.2 (Schubert *et al*., 2016) to enforce a minimum sequencing quality threshold of 20 for individual bases, trim reads past the first low-quality base, trim Ns at the end of reads, and keep only reads of a minimum length of 20 bases. For reference genome mapping, we used collapsed forward and reverse files to increase quality scores and allow more sensitive mapping to references. We evaluated data quality control using FastQC (Andrews, 2010) before and after the trimming step, later aggregating results across all samples into a single report using MultiQC (Ewels *et al.*, 2016). We performed all bioinformatics analyses (described in Supporting Information, Supplemental File 4) in the Smithsonian Institution High-Performance Computing Cluster (https://doi.org/10.25572/SIHPC).

Assessment of Exogeneous Content

We define exogenous DNA as genetic material not assigned to the target species of lorisiforms, such as any post-mortem environmental contaminants (e.g., human, fungi, bacteria, and other microorganisms) or potential parasite species present in the tissue of the animal while in life. We performed two analyses to evaluate 1) the levels of human contamination and 2) the relative abundance of genetic material derived from nonchordate and nonplant exogenous content present in libraries sequenced from museum samples. First, we monitored the fraction of high-quality hits mapped exclusively to the human mitogenome reference using BWA aln, relative to the total number of reads mapped to the closest lorisiform species mitochondrial reference using BWA mem. The underlying assumption behind our strategy is that truly human reads will map with high precision to the *Homo sapiens* mitochondrial reference using BWA aln (an algorithm with limited performance when divergence from the reference genome is high; see Supplemental File 1 for more details), whereas reads from highly conserved regions across primates will map to both references with equal quality.

Second, we evaluated the levels of exogenous content through a metagenomic approach using Kraken2 (Wood *et al*., 2019), a fast k-mer-based approach for assigning DNA sequences to nested taxonomic categories. First, we mapped all trimmed and quality-filtered reads to the *Otolemur crassicaudatus* draft assembly generated by the DNA Zoo Team (https://www.dnazoo.org). We then calculated the ratio between the number of reads mapped and the total number of reads sequenced used this mapping success per sample as a proxy for the level of endogenous content. To estimate the level of library redundancy, we calculated the proportion of unique fragments mapped to the closest reference genome. In the metagenomics analyses,

we used only the unmapped reads as input file. We ran Kraken2 through the command line using the PlusPFP database (which includes indexed reference sequences for archaea, bacteria, viruses, plasmids, humans, protozoa, fungi, and plants). We imported individual samples' report to into the interactive Pavian Shiny app in R (Breitwieser & Salzberg, 2020) to filter out identifications based on less than 10 hits and using a minimum threshold of 0.001%, and export tables and Sankey flow diagrams for each sample. In our exploratory metagenomic analysis, we focus only on hits classified into one of the four high-level taxonomic categories: bacteria, viruses, archaea, and eukaryotes (excluding Chordata and Viridiplantae) domains. To quantify the composition and similarity of the community of classified exogenous content, we calculated the alpha (Simpson's complementary diversity index) and beta diversity (Bray-Curtis dissimilarity index) using information about the number of reads classified to the level of species for each sample. Both indices range from 0 to 1. The Simpson diversity index indicates the community richness in each sample, with values closer to one indicating the highest diversity in the pool of samples. The Bray-Curtis dissimilarity index indicates how similar two samples are to one another, with 0 indicating a complete overlap in species composition. To calculate and analyze these statistics, we used the "Phyloseq" (McMurdie & Holmes, 2013) and "vegan" (Oksanen *et al*., 2022) R packages.

Mitochondrial Genome Reference Mapping and Base Calling

We then evaluated the impact of reference bias, that is, the impact of evolutionary divergence between the focal species sequenced and the one serving as a reference genome, on the final mapping success and quality. Specifically, we compared the performance of the mapping step using four reference primate species that varied in levels of divergence from our focal lorisiform species: *Homo sapiens*, *Perodicticus potto*, *Galago senegalensis*, and *Otolemur crassicaudatus* (Supplementary File 1; Table S1). We focused on the mitochondrial genome, as this organelle is widely employed as a proxy for evolutionary divergence at the species level and is a useful starting point for future taxonomic studies (Avise, 1991; Rubinoff & Holland, 2005; Ballard & Rand, 2005; Groves, 2001).

We downloaded reference mitogenomes from NCBI (accession numbers are provided in Supplementary File 1; Table S1), and indexed them before mapping using samtools (Li 2011). To map raw reads to references, we used the Burrows-Wheeler Aligner (BWA V.0.7.17) MEM algorithm, which performs well with fragments >75 bp and handles deep divergences robustly; however, its performance is impaired by read lengths shorter than 75 bp (see Supporting Information for details of the comparison of mapping algorithms; Fig. S1). After mapping, we sorted all BAM files and discarded unmapped reads (flag -F 4) for downstream analyses. BWA has a limited capacity to map reads around short indels, which can interfere with variant calling and introduce errors in downstream analysis. Therefore, we used GATK (Van der Auwera & O'Connor, 2020) to locally realign the reads using a three-step approach. For each reference genome, we created a local dictionary using picard

tools (Broad Institute, 2019). We added read groups to our sorted bam files using samtools (Li, 2011), created and realigned read intervals using GATK.

We then performed quality control and filtering to remove PCR duplicates and reads that mapped to ambiguous positions. We excluded reads shorter than 30 bp and reads with a mapping quality lower than 20, which have a higher chance of mapping to nonspecific genomic regions. To calculate the depth coverage per locus in the resulting sorted bam files, we used samtools *depth* with the–aa option to report absolutely all positions. We performed consensus calling while retaining positions with a minimum coverage of 2x, removing indel-containing calls, retaining base calls, and calling the most common base. We then calculated the number of reads mapped and the mean and standard deviation of coverage across all the loci per sample. Finally, we used *t*-tests to evaluate whether the number of mapped fragments differed when using the closest lorisiform species as a reference relative to the human genome. We visualized the improvement in mapping performance by plotting the divergence to the reference against the increase in coverage (ratio between the mean coverage when mapped to the closest reference and that using the human mitogenome).

Phylogenetic Analyses

After sequence mapping and consensus calling, we inferred evolutionary relationships between our focal lorisiform samples based on complete mitogenomes. To this end, we used only the consensus files obtained by mapping the raw reads of each sample to the closest reference mitogenome available for that species (Table S1). We merged all final consensus FASTA files into a multi-alignment. We then generated a final alignment including our newly generated sequences, the three lorisiforms used as references, two Asian lorisids (*Loris tardigradus* and *Nycticebus cougang*, and one lemuriform species as an outgroup (*Microcebus murinus*; see Table S1 for accession numbers) using Multiple Alignment using Fast Fourier Transform (MAFFT) as implemented in Geneious Prime (V.2021.2.2). We then inspected the alignment to check for the presence of stop codons using annotations from the closest mitochondrial reference and the vertebrate mitochondrial translation code. Based on this alignment, we performed phylogenetic inference by incorporating variant and invariant sites under the Maximum Likelihood. To this end, we ran RaxML-HPC v. 8.2.12 (Stamatakis, 2014) employing the GTRCAT model of nucleotide evolution and estimating node support based on 1,000 bootstraps. We focused our analysis on two phylogenetic scales: a deeper divergence between genera, and a shallower scale within specimens of western dwarf galagos assigned to the genus *Galagoides*.

Ethical Note

We only collected tissues from preserved museum specimens. After obtaining official approval from the collection authorities (curators and collection managers) some of the authors (A.P., M.B., and L.P.) sampled the museum specimens and shipped

the tissues in individual tubes to the Smithsonian National Museum of Natural History or American Museum of Natural History for DNA extraction at historical DNA facilities. The authors declare no potential conflict of interest.

Data Availability Statement Raw data generated in this study are deposited in the National Institute of Health Sequence Read Archive (SRA) under BioProject number SUB14275283.

Results

Performance of DNA Extraction Protocols, Sequencing, and Mapping Strategies

Both protocols recovered DNA from historical museum samples collected during a 100-year period spanning 1896–1987 (N = 238). All negative control extractions (N = 20) had undetectable DNA concentrations (i.e., qubit reads below a measurable range). The oldest specimens we obtained DNA was collected in 1896 (Protocol 1) and 1912 (Protocol 2). We found a significant correlation between sample age and the total amount of recovered DNA for Protocol 1 (Pearson's correlation coefficient = 0.42, $p < 0.001$; df = 147) but not for Protocol 2 (Pearson's correlation coefficient = 0.07, $p = 0.482$; df = 89; Fig. 2a). On average, Protocol 1 also yielded higher final DNA concentrations for samples of all ages (t-test = 8.29 , $p < 0.001$; df = 236; Fig. 2b; Protocol 1: min = 0.102; mean = 76.4; max = 330 ng/μL; SD = 62.18; N = 147) than Protocol 2 (min = 0.034; mean = 8.936; max = 60 ng/μL; SD = 11.07; N = 89).

We obtained a mean of 9.6 million reads for each sequenced library (min = 0.1 $*10^6$, max = 23.2 $*10^6$, SD = 4.11; N = 96), and the mean fragment length of those extracted using Protocol 1 (min = 54 bp; mean = 97 bp; max = 133 bp; SD = 36.85, N = 55) was significantly smaller than that of those extracted using Protocol 2 (min = 52 bp; mean = 111 bp; max = 151 bp; SD = 37.45; N = 41, t-test = 2.85 ; $p < 0.001$; df = 94; Fig. 2c), indicating that the modified phenol-chloroform protocol with MinElute columns retained a greater proportion of small fragments overrepresented in historical samples, which are likely discarded by standard kits. After removing adaptors, the mean fragment length of our libraries was 103 bp (min = 52; max = 151; SD = 20; N = 96). PCR duplicate levels ranged 0.03–0.46 (mean = 0.13; SD = 0.05; N = 96). The mean GC% of the sequenced fragments was 44 (min = 40; max = 52; SD = 02; N = 96).

The mapping performance improved with lower phylogenetic divergence from the reference genome. Using the closest species of lorisiform as a reference outperformed the human mitogenome in terms of number of reads mapped (Fig. 3a), as well as the mean coverage across all loci, which increased by 5- to 20-fold, depending on the taxon (Fig. 3b).

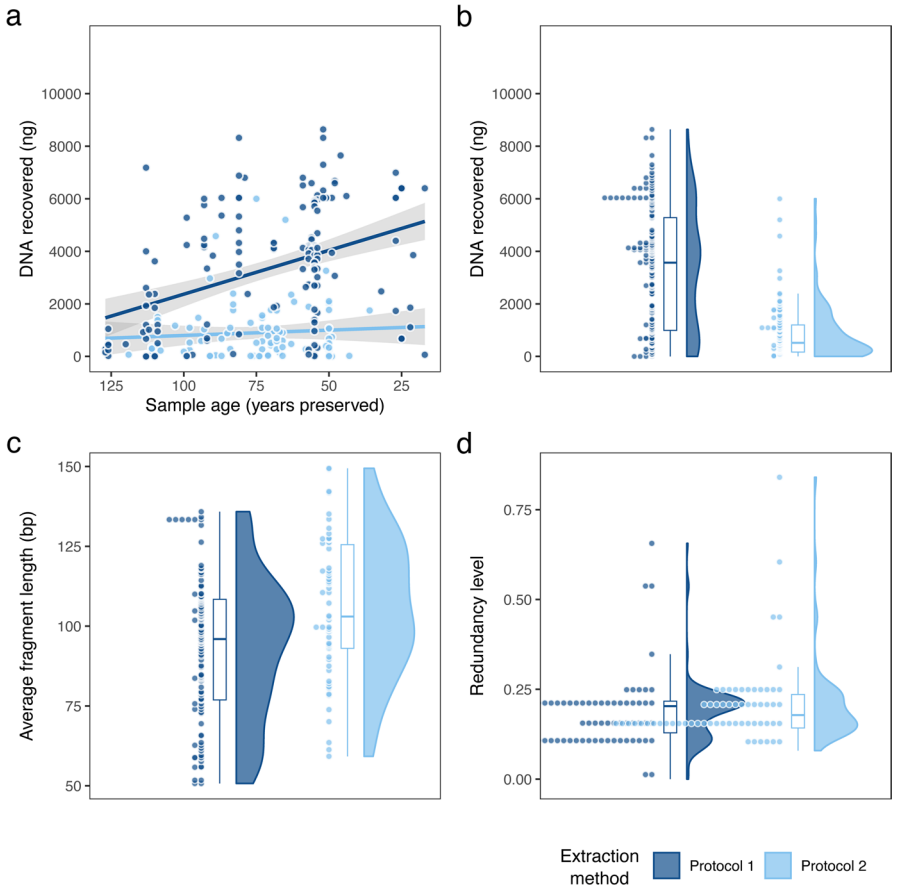

Fig. 2 Comparison of the performance of two DNA extraction protocols using lorisiform museum specimens. (**a**) Relationship between sample age and total DNA template recovered using the two protocols compared in this study. Boxplots and distributions of final DNA recovered (**b**) for each protocol, mean insert size after adapter removal (**c**), and redundancy levels (**d**) in libraries sequenced using the different protocols. In the boxplots, the box indicate the interquartile range (IQR) from first (Q1) to the third (Q3) quartile, and internal line represents the median. Whiskers indicate the variability outside the interquartile range and are calculated as Q1/Q3 ± 1.5 * IQR. Points outside the whisker ranges can be considered as outliers.

Assessment of Exogenous Content

Both protocols resulted in a mean of 88% of endogenous content (min = 49.3%; max = 99.8%; *SD* = 12), and only ten samples had less than 80% of the reads mapped to the closest reference (Fig. 4). Our analyses of mapping quality estimated negligible levels of human contamination for both protocols (Figs. 4a and S1), with only one exception out of the 96 libraries analyzed (RMCA 25984, which had 3.8 times more reads uniquely mapped to humans than those mapped to both references with equal quality). Our metagenomics analysis failed to classify a mean of 11% of total reads

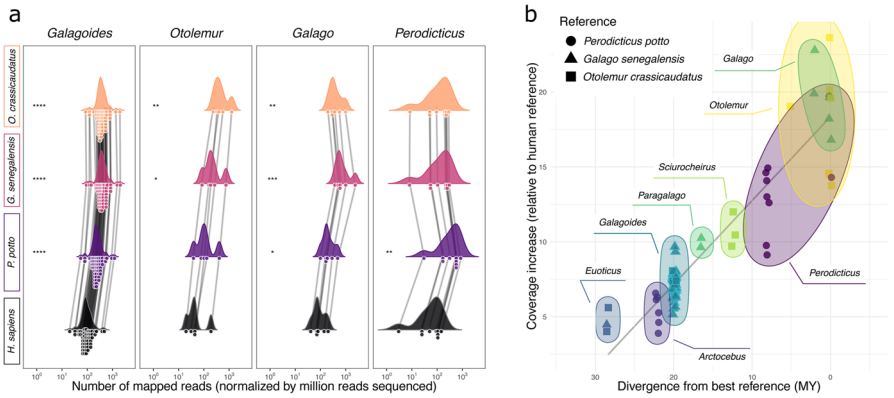

Fig. 3 Mapping success using different reference mitogenomes. (**a**) Vertical panels subset the results by genera, and the colored distributions show the four different mitochondrial references used (from top to bottom: *Otolemur crassicaudatus*, *Galago senegalensis*, *Perodicticus potto*, and *Homo sapiens*). Each point represents the normalized number of reads mapped to the different reference genomes for a given sample. Gray lines connect the results of individual samples using the four different references. Asterisks indicate significance level in the t-test comparing the mean number of mapped reads when using one of the Lorisiform species versus the human reference mitogenome (no asterisk: $p > 0.05$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; ****$p \leq 0.0001$). (**b**) Relationship between the relative increase in coverage (ratio between mean coverage when using the closest mitochondrial reference relative to when using the human mitochondrial reference; y axis) and the divergence to the closest mitochondrial reference (x axis).

into any taxonomic category (min = 0.2%; max = 49%; *SD* = 0.8), and only classified a minimal fraction (min = 0.01%, mean = 0.6%, max = 3.2%, *SD* = 0.46) into any taxonomic rank (Fig. 4a). The relative abundance of reads classified within these groups varied among samples (Fig. 4b) and protocols. Protocol 2 exhibited, on average, a higher amount of exogenous content classified as Bacteria than Protocol 1 (Fig. S3). Protocol 1 showed higher amounts of Phylum Baciliota, Bacteroidota, Euglenozoa, Uroviricota, Euryarchaeota, and Thermoproteota, whereas Protocol 2 showed higher proportions of Pseudomonadota and Actinomycetota (Fig. S3).

The analysis of community composition showed high species richness, with no significant differences between protocols (Fig. 4c). Our analyses revealed moderate to high dissimilarity between the community compositions of samples extracted using each protocol, and more similarity among libraries generated from Protocol 1 than those obtained using Protocol 2 (Fig. 4d). Only one sample showed extremely low alpha diversity (USNM 377274, Simpson index = 0.09, with *Proteus mirabilis* dominating the classified reads), whereas the majority had much higher species richness in the community of identified taxa (mean = 0.89; max for RMCA 17905, Simpson index = 0.99; Fig. S4). Interestingly, our analyses suggest that the metagenomic community was dominated by a single organism in only a few samples. For instance, the bacteria *Proteus mirabilis* corresponded to 92.1% of the classified reads of sample USNM 377274, and *Staphylococcus saprophyticus* corresponded to 80.7% of those of sample RMCA 8732, but both organisms showed extremely low (if any) amounts in the remaining

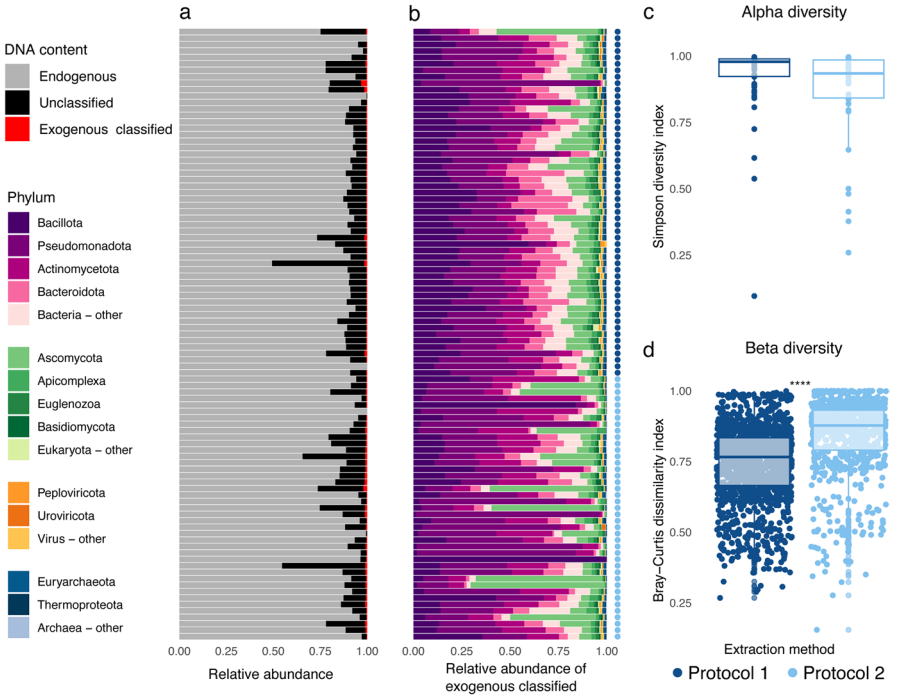

Fig. 4 Metagenomic and diversity analyses for Lorisiform museum specimens. Relative abundance of total reads (**a**), with endogenous (mapped to the closest reference), unmapped and unclassified, and exogenous (unmapped reads later classified using Kraken2 at domain level). Horizontal bar graphs represents the proportion of each content by sample. Samples are sorted by age (oldest on top) and by protocol (colored blue circles on the right). Detail of relative abundance of the exogenous content is presented at Phylum level (**b**). The metagenomic community diversity (**c**) of the samples extracted by using different protocols and the similarity of the two communities (**d**) was estimated by calculating the alpha (for each sample) and beta diversity (pairwise) and presented as boxplots by protocol. In the boxplots, the box indicate the interquartile range (IQR) from first (Q1) to the third (Q3) quartile, and internal line represents the median. Whiskers indicate the variability outside the interquartile range, and are calculated as Q1/Q3 ± 1.5 * IQR. Points outside the wisker ranges can be considered as outliers. Significance levels are indicated by asterisks (no asterisk: $p > 0.05$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; ****$p \leq 0.0001$).

samples (Figure S5). In contrast, our analyses identified *Staphylococcus aureus*, *Escherichia coli*, and *Aspergillus luchuensis* in 85 of the 96 samples analyzed at various levels (Figure S4). Our analysis also indicated the presence of the hemosporid apicomplexan of the genus *Plasmodium*, trypanosomatid euglenozoan of the genus *Leishmania*, cholera bacterium of the genus *Vibrio*, syphilis-causing bacterium of the genus *Treponema,* and penicillin-resistant bacteria of the genus *Pseudomonas.* Although our analysis assigned different number of reads to these organisms (Figure S4), all samples exhibited minimal proportions of exogenous classified reads (Fig. 4a).

Phylogenetic Analyses

Our phylogenetic analysis inferred the monophyly of galagids and the known phylogenetic relationship between all galagid genera: *Euoticus* as the sister taxon to the remaining genera, followed by *Galagoides*, *Sciurocheirus* as the sister taxon to *Otolemur,* whereas *Paragalago* as the sister taxon to *Galago* (Fig. 5a). Our analysis also recovered *Arctocebus* as sister to *Perodicticus* and *Loris* as sister to *Nycticebus*; however, the support for the monophyly of the Lorisidae family was low (bootstrap = 48).

Finally, our phylogenetic results suggest interesting biogeographic patterns in the western dwarf galagos. Remarkably, *Galagoides* samples assigned to distinct taxa were not geographically partitioned. Instead, samples assigned to distinct taxa formed two major sympatric clades (I and II in Fig. 5) that showed levels of genetic divergence comparable to those observed between pairs of other Galagidae sister genera. Within each of these two sympatric clades, samples were clustered according to geography. Clade I includes a highly divergent lineage from western Africa, which is sister to the remaining lineages in this clade. These lineages include a west-central lineage and a central lineage. Clade II includes a highly divergent west-central lineage sister to two lineages: one consisting of two western-central lineages and a western lineage and the other consisting of a central lineage. An expanded phylogeny showing all 96 samples and the reference mitogenomes used in this study can be found in the Supporting Information (Fig. S6).

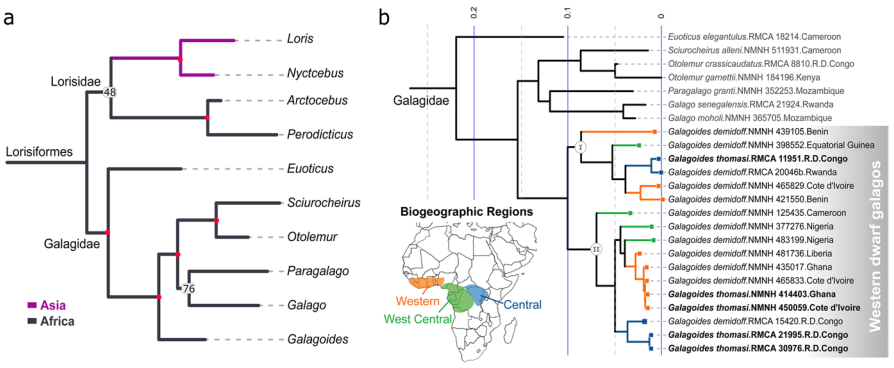

Fig. 5 Phylogenetic relationships between lorisiforms based on museum samples, with special attention to western dwarf galagos. (**a**) Maximum Likelihood best tree using complete mitogenomic data showing the relationships between all currently recognized Lorisiform genera using *Microcebus murinus* as outgroup. Values at nodes show bootstrap support value when different from 100 (red dots). (**b**) Phylogeographic contextualization of western dwarf galagos (genus *Galagoides*) samples used in this study showing a detail of the Maximum Likelihood best tree using complete mitogenomic data. Branch colors correspond to biogeographic regions highlighted in the map. Tip labels in bold indicate museum specimens labeled as "*Galagoides thomasi.*" Clades labeled as "I" and "II" indicate the two major clades within *Galagoides.*

Discussion

Based on 238 specimens of 19 lorisiform species deposited in four natural history collections on two continents, our investigation showed that younger samples and a phenol-chloroform extraction followed by MinElute column purification step resulted in higher final DNA amounts. We also found that using the human mitogenome as a reference results in lower mapping coverage, but coverage increases with higher phylogenetic proximity to the lorisiform species used as reference. Our metagenomic investigation revealed the potential of museum specimens to detect the presence of exogenous bacterial contaminants likely associated with long-term specimen preservation as well as potential parasites through "genomic bycatch" (Holmes & Davis-Rabosky, 2018; Zacho *et al*., 2021). Our phylogenetic analyses based on mitogenomes also confirmed previously established phylogenetic relationships between all Galagidae genera but found low support for the monophyly of African and Asian Lorisidae genera. Lastly, our results provide insights into the biogeographic patterns of western dwarf galagos (genus *Galagoides*), while revealing fundamental inconsistencies in the taxonomic labeling of museum specimens.

Effects of Sample Age, DNA Extraction Protocol, and Bioinformatic Strategy

Our efforts to obtain DNA from historical primate species were very successful, yielding DNA from specimens from 35 to 126 years after their collection at field sites across Sub-Saharan Africa. Overall, more recently preserved specimens yielded higher DNA concentrations. We found that a stringent approach that includes a bleach wash, followed by phenol-chloroform extraction and MinElute column DNA purification steps (Protocol 1), outperformed the commercially available DNEasy Blood and Tissue Qiagen kit (Protocol 2) in terms of DNA yield regardless of sample age. Protocol 2 produced similar concentrations across samples regardless of sample age, suggesting some limitation in the extraction kit. Moreover, Protocol 1 retained smaller fragments (>70 bp) than Protocol 2, consistent with the differences in membrane permeability of the purification columns used in each of the protocols.

Our shotgun sequencing approach assessed the level of endogenous content and library complexity. Both protocols resulted in similar mapping success, library redundancy, and exogenous species richness. The lower levels of beta diversity observed in the libraries obtained from extractions using Protocol 1 suggest that the exogenous content of extractions using this protocol is more homogeneous than those obtained using Protocol 2. This could be explained by the prewash step adopted in Protocol 1, which we used in an effort to minimize contamination from any DNA source in the external parts of the sample before DNA extraction (Korlević *et al.*, 2015), and even increase the amount of exogenous DNA recovered (Hajdinjak *et al.*, 2018). However, we found no difference in the levels of exogenous DNA recovered between samples with and without the bleach pre-wash. Overall, our results suggest that while a simple commercial kit can recover consistent DNA concentrations from historical samples, a more laborious protocol designed specifically for archival DNA can greatly increase extraction yields while maximizing the

retention of shorter fragments that characterize historical samples. These smaller fragments can contribute to increasing library complexity, as genetic data for the same region can be obtained using multiple independent DNA molecules, rather than just PCR products emerging from library amplification.

These results indicate that the modified phenol-chloroform protocol designed specifically for fragmented archival DNA has superior performance relative to standard commercial kits for DNA extraction in terms of the final amount of DNA recovered, thus reinforcing the results of previous experiments (Campos & Gilbert, 2019). Although Protocol 2 recovered lower amounts of DNA from the historical samples, it generated sufficient DNA material for genetic research. One limitation of our study is that we extracted each sample using only one protocol, so we could not compare the extraction methods directly. Nevertheless, our results are similar to those of other studies that compared different DNA extraction protocols more systematically (Hawkins, Bailey *et al*., 2022a, Hawkins, Flores *et al*., 2022b; McDonough *et al*., 2018).

Given the limited genomic resources currently available for the species we studied, we examined the impact of reference bias by comparing mapping performance and the retrieval of historical DNA sequences using reference genomes of varying phylogenetic divergence to the focal species. We found that increased divergence from the reference significantly reduced the coverage of the final consensus sequences. Relative to using the human mitogenome (a highly divergent species to all lorisiforms) as a reference to assemble the genome of our focal lorisiform species, we observed a fivefold increase in coverage when using a reference from the sister family or genus, tenfold increase when using a reference from the same family or genus, and a 20-fold increase when using a reference from the same species. In addition to higher coverage, closer reference genomes also increased the total number of mapped reads. These results indicate that the reference sequence used for read-mapping is a major constraint in bioinformatics processing of fragmented DNA from museum samples. More specifically, the levels of evolutionary divergence from the species serving as a reference genome has a disproportionate effect on the final read coverage and length of the historical DNA. We recovered a high number of mapped reads with depths that are appropriate for most evolutionary applications (>10x) when incorporating a reference from the same genus or family as the focal species. Nevertheless, the sequences generated using these more distant references provided key insights into the evolution and distribution of African galagids, attesting to the value of museum specimens, even when genomic resources are limited. Future studies incorporating archival DNA from other primate species will benefit from developing genomic resources for the focal taxon (e.g., high coverage whole genome from high-quality samples, long-read assemblies), which may increase the project costs and complexity. The rapid advances in generating publicly available primate genomic resources will be crucial to this effort (Kuderna *et al.*, 2023).

Metagenomics Reveals Likely Environmental Contaminants and Parasites

Primate collections are among the most widely used mammal collections, and researchers often handle specimens without protective gear (e.g., gloves, masks, and hair nets). Because of their close phylogenetic relationship to humans, it is important to investigate the degree of potential human contamination in libraries sequenced from primate study skins. Our human contamination analysis of mapped reads indicated negligible levels of human and other exogenous contamination in both protocols. The only sample with elevated proportions of reads with higher mapping score to human reference than to the closest reference (3.8-fold for RMCA 25984) also showed the lowest mapping success ratio (0.54) and small mean fragment length (59 bp). Regardless, the 8.5 million reads mapped to the closest reference were sufficient to place this sample among others of the same species that showed negligible levels of human contamination (0.0077 for RMCA 18873 and 0 for RMCA 14674).

Our metagenomics analysis suggests that the vast majority of the sequenced reads generated using a shotgun approach on museum samples corresponded to DNA from the focal species. Although 86 of 96 samples analyzed had more than 80% of the reads classified as endogenous content and more than 11% not classified because of short fragment size, a minimal proportion (0.06% on average) of reads were taxonomically assigned to archaea, viruses, bacteria, or nonplant and nonchordate eukaryotes. The relative abundance of this small fraction of exogenous classified content did not show any relationship with sample age or a specific extraction protocol. Our analysis suggested the presence of two categories of exogenous content. First, our analyses classified reads into to several modern strains of free-ranging, multidrug-resistant bacteria across all samples (Fig. S5). Because none of our negative control extracts yielded detectable DNA using a high sensitivity fluorometer test, it is unlikely that these nontargeted sequences originated from reagent contamination in the wet laboratory. Instead, our results suggest that these organisms can likely adhere and live in the tissue of study skins amidst the chemicals employed during preservation and through care practices inside the museum cabinets. However, our study did not try to determine whether these organisms were active or possessed antimicrobial resistance. Future investigations might reveal the origins, variations, and potential consequences of such contaminants in natural history collections.

Another interesting and somewhat unexpected group of microorganisms identified through our metagenomic approach are potential parasite species known to cause blood-borne diseases in primates, such as malaria (*Plasmodium*) and leishmaniasis (*Leishmania*), and foodborne diseases, such as cholera (*Vibrio*). Despite recovering DNA sequences that matched these potential parasites, we could not determine whether the primate specimens sampled were sick in life. Moreover, it is unclear whether these organisms were acquired after specimen preservation or while the animals were still alive. To distinguish between these scenarios, future studies should compare the signatures of DNA decay between reads assigned to focal species and potential parasites. It also is possible that the metagenomics approach we applied in our exploratory analysis failed to assign unmapped reads to the true taxa, and these classifications are false positives. Future studies should compare Kraken2 outputs with those from other bioinformatic tools more appropriate for degraded

samples (Velsko *et al* 2018), validating classification results through a backward approach via BLAST (Basic Local Alignment Search Tool; Altschul *et al.*, 1990) to assess identity scores between the reads flagged by metagenomics and the available genetic information for these organisms in public repositories.

The detection of parasitic DNA in museum specimens has implications for wild primate and human health. Nonhuman primates are important reservoirs for zoonotic diseases and other opportunistic pathogenic organisms (Goméz *et al.*, 2013; Kotait *et al.*, 2019; Kebede *et al.*, 2020). Because of their close evolutionary relationships, humans can be at risk of infection from primate-borne parasites. However, little is known about parasites in most wild primate populations (Gillespie *et al.*, 2008; Cunningham *et al.*, 2017; Solórzano-García and Pérez-Ponce de León, 2018). In the case of galagos and lorises, two studies have suggested that disease transmission to humans can occur during contact with blood and other bodily fluids through bushmeat consumption or pet trade (Fominka *et al.*, 2021; Svensson *et al.*, 2021). Broadly, our results support the idea that museum specimens can provide insights into the presence of parasites in natural primate populations collected from distinct localities and at different points in time (Thompson *et al.*, 2021). In turn, this information might inform primate conservation and assessment of disease transmission risk to regional and global human communities. We hope that our study provides a basis for future research.

Based on results from our shotgun strategy coupled with an exploratory metagenomic approach, we provide recommendations for researchers interested in incorporating museum specimens in their genetic studies. Conducting shotgun screening runs of all museum samples available at lower sequencing effort can be a strategic approach to determine the levels of exogenous contaminant DNA relative to the target species DNA across samples. These preliminary runs can be used to select samples with minimal contamination for subsequent high-coverage sequencing or, alternatively, elaborate target capture strategies to maximize the enrichment of sequences from target organisms (e.g., mitochondrial or UCE capture). Genetic analyses of archival DNA also can benefit from bioinformatic detection and removal of microorganism DNA contamination, thus improving genome assembly of the focal organism. Although increasing total costs, this strategy can prevent the sequencing of samples overrun by bacterial and other environmental contaminants.

Phylogenetic Patterns

Our study also inferred the phylogenetic relationship of lorisiform samples based on mitogenomes. Overall, our results confirm that DNA sequences obtained from museum specimens are reliable indicators of primate evolutionary history above and below species level. We found strong support for the most current hypothesis of phylogenetic relationships among the six Galagidae genera (*Euoticus,* (*Galagoides,* (*Galago, Paragalago*), (*Otolemur, Sciurocheirus*))) and the sister relationship between the two African (*Arctocebus, Perodicticus*) and two Asian Lorisidae genera (*Loris, Nycticebus*; Pozzi *et al.*, 2015). Thus, our results support the presence of two distinct lorisiform radiations on the African

continent (Munds *et al*., 2018; Pozzi, Roos *et al*., 2020b). However, our analyses recovered very low support for the monophyly of African and Asian lorisids, mirroring the results from other studies that also relied on mitochondrial data (Finstermeier *et al*., 2013; Pozzi *et al*., 2014b). This finding might reflect a limited phylogenetic signal in the mitochondria that fails to recapitulate the evolutionary history of the lorisiform radiation. Notably, whereas we included variable and invariable sites from the complete mitochondrial genome, all segregating sites represent a single locus. Pozzi *et al*. (2015) suggested that this low support might also reflect the rapid and early divergence between the Asian and African lineages (37.85 MY, shortly after the split from Galagids at around 39.7 MY), leading to incomplete allele sorting between these major lineages (Pozzi *et al*., 2015, Pozzi, Roos *et al*., 2020b). Future studies aiming at elucidating the phylogenetic relationship and time of divergence between African and Asian lorisids will benefit from incorporating genomic-level data for multiple strepsirrhine taxa, including lemur species.

In addition to information on genus and family level relationships, our results provide new insights into the relationships between recently divergent species and the taxonomic assignment of closely related populations. We focus more on the two broadly distributed species of western dwarf galagos (*Galagoides demidoff* and *G. thomasi*). We found that specimens morphologically assigned or otherwise labeled as *Galagoides demidoff* and *G. thomasi* do not correspond to monophyletic units. Specifically, specimens attributed to each of these taxa were often grouped together, whereas samples corresponding to the same taxon often were not found to be closely related. Instead, the samples were grouped into two broadly sympatric lineages that co-occurred along the lowland forests of western, western-central, and central Africa (clades I and II in Fig. 5b). The level of divergence between clades I and II (as indicated by branch lengths) was within the range observed between sister galagid genera (e.g., *Sciurocheirus* and *Otolemur*, *Galago* and *Paragalago*). These results suggest that specimens assigned to species of *Galagoides* are either filed under an outdated taxonomy or were misidentified in the field, and clades I and II correspond to the currently recognized species *G. demidoff* and *G. thomasi*. Alternatively, the genus consists of several cryptic lineages restrained to smaller geographic ranges, concurring with the perception of unclear species boundaries in the genus. However, because our study evaluated only mitochondrial DNA, this pattern could also reflect ancient introgression events. To test these hypotheses and determine whether these lineages correspond to previously recognized subspecies or even cryptic unrecognized diversity, future studies should delineate lineage boundaries using nuclear and mitochondrial markers and integrate phenotypic and genomic data to evaluate whether genetic patterns correspond to diagnosable phenotypic variation within the genus. Our results confirm previous assertions of taxonomic confusion and unclear species limits in *Galagoides* (Groves, 2001; Grubb *et al*., 2003; Masters *et al*., 2017) while confirming the enormous potential of historical DNA to fill these knowledge gaps.

Conclusions

Currently, human activity is causing biodiversity crises. Primate populations are rapidly decreasing, leading to species extinction worldwide (Ceballos *et al.*, 2015; Estrada *et al.*, 2017; Davis *et al.*, 2018). The vast majority of lorisiform systematic studies have relied on museum collections, but cryptic species may be difficult to identify based on morphology alone, especially for nocturnal taxa. Our study demonstrates the critical role of museum collections by bridging specimen-based research and cutting-edge genomic and bioinformatics techniques in our understanding of the evolution of lorisiforms. We cannot predict the full spectrum of information that future generations will be able to retrieve from today's specimens. This realization comes with great responsibility. Specimen manipulation and destructive sampling require careful consideration of their potential long-lasting effects (Raxworth & Smith, 2021). Our results show that protocols specially designed for degraded samples allow higher DNA yield and retention of smaller fragments, improving library complexity. We found that the human reference mitogenome is not appropriate for use as a reference in lorisiform genomic studies. More efforts should be devoted to generating high-quality reference genomes for species of the genera *Euoticus, Galagoides, Paragalago, Sciurocheirus,* and *Arctocebus*, for which no genome-level data are currently available. Our analyses indicate minimal amounts of exogenous content in shotgun sequencing of museum specimens. Results of our exploratory metagenomic investigation attest to the promise of archival DNA as an indicator of parasite presence in wild specimens represented in natural history collections. Lastly, our study demonstrates how DNA obtained from museum specimens can fill important taxonomic and geographic gaps required to investigate deep phylogenetic relationships and answer long-lasting biogeographic questions in broadly distributed cryptic primate species.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10764-024-00429-3.

Acknowledgments The authors thank Michelle Sauther, Frank Cuozzo, and Jack Dalton for the invitation to contribute to the AABA Bushbaby symposium (Denver, CO, 2022), as well as Joanna M. Setchell, Ivan Prates, Christina Bergey, Eva Wikberg, Markus Hamilton and two anonymous reviewers for their insightful comments on an earlier version of this manuscript. A.P. thanks Rayna Bell (California Academy), Melissa Hawkins (Smithsonian National Museum of Natural History), Ivan Prates (U. Michigan and Lund University), and Lilian Parker (Smithsonian's National Zoo and U. Oklahoma) for their thoughtful conversations on this topic over the years. The authors also thank all the curators and collection managers (Darrin Lunde, Megan Viera, Michael McGowen, Melissa Hawkins, Neil Duncan, Eleanor Hoeger, Emmanuel Gilissen, and Jean-Marc Herpers) who facilitated the sampling of specimens used in this study; and the Smithsonian Laboratory of Analytical Biology and Vertebrate Zoology (Jeffrey Hunt, Katherine Murphy, and Mary Faith Flores) and Hydra High-Performance Computing (Rebecca Dikow and Matthew Kwenskin) staff for assistance with laboratory and bioinformatics logistics respectively. A.P. was funded by the American Society of Mammalogy (ASM), Society for Systematic Biology (SSB), Systematics Association, Linnean Society of London, and National Science Foundation of the United States (NSF; Award No. BSC-2120691); A.P., E.P., and L.K. were funded by the Smithsonian Institution Graham Funds in Human Origins. M.E.B. and L.P. were funded by the National Science Foundation of the United States (Awards No. BCS-1926105/1926215). A.P. and L.P. were also funded by the University of Texas at San Antonio (UTSA).

Author Contributions Conceptualization: A.P., M.E.B.., L.P.; Sample acquisition: A.P, M.E.B, L.P.; Funding acquisition: A.P, M.E.B, L.K., L.P.; Methodology: all authors; Analysis: A.P.; Writing –original draft: A.P; Writing –review and editing: all authors.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://www.sciencedirect.com/science/article/abs/pii/S0022283605803602?via%3Dihub

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ang, A., Roesma, D. I., Nijman, V., Meier, R., Srivathsan, A., & Rizaldi. (2020). Faecal DNA to the rescue: Shotgun sequencing of non-invasive samples reveals two subspecies of Southeast Asian primates to be Critically Endangered species. *Scientific Reports, 10*(1), 9396. https://doi.org/10.1038/s41598-020-66007-8

Avise, J. (1991). Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annual Review of Genetics, 25*, 45–69.

Aylward, M. L., Sullivan, A. P., Perry, G. H., Johnson, S. E., & Louis, E. E. (2018). An environmental DNA sampling method for aye-ayes from their feeding traces. *Ecology and Evolution, 8*(18), 9229–9240. https://doi.org/10.1002/ece3.4341

Ballard, J. W. O., & Rand, D. M. (2005). The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology, Evolution, and Systematics, 36*, 621–642.

Blair, M. E., Cao, G. T. H., López-Nandam, E. H., Veronese-Paniagua, D. A., Birchette, M. G., Kenyon, M., et al. (2023). Molecular Phylogenetic Relationships and Unveiling Novel Genetic Diversity among Slow and Pygmy Lorises, including Resurrection of Xanthonycticebus intermedius. *Genèses, 14*(3), 643. https://doi.org/10.3390/genes14030643

Bock, S., & Quaisser, C. (2019). Tear and Crumble: Deterioration Processes in Skins and Hides in Mammal Collections. In Collection Forum (Vol. 33, No. 1, pp. 36-54). Soc. for the Pres. of Natural History Collections. https://doi.org/10.14351/0831-4985-33.1.36

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology, 13*(11), 3261–3273. https://doi.org/10.1111/j.1365-294X.2004.02346.x

Breitwieser, F. P., & Salzberg, S. L. (2020). Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics, 36*(4), 1303–1304. https://doi.org/10.1093/bioinformatics/btz715

Broad Institute. (2019). Picard toolkit: A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. *Broad Institute, Picard toolkit GitHub repository*. https://github.com/broadinstitute/picard. Accessed 10 Oct 2022

Brown, R. M., Weghorst, J. A., Olson, K. V., Duya, M. R. M., Barley, A. J., Duya, M. V., et al. (2014). Conservation genetics of the Philippine tarsier: cryptic genetic variation restructures conservation priorities for an island archipelago primate. *Plos One, 9*(8), e104340. https://doi.org/10.1371/journal.pone.0104340

Buerki, S., & Baker, W. J. (2016). Collections-based research in the genomic era. *Biological Journal of the Linnean Society, 117*(1), 5–10. https://doi.org/10.1111/bij.12721

Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2015). The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution, 79*, 35–44. https://doi.org/10.1016/j.jhevol.2014.10.015

Campos, P. F., & Gilbert, M. T. P. (2019). DNA Extraction from Keratin and Chitin. *Methods in Molecular Biology, 1963*, 57–63. https://doi.org/10.1007/978-1-4939-9176-1_7

Card, D. C., Shapiro, B., Giribet, G., Moritz, C., & Edwards, S. V. (2021). Museum Genomics. *Annual Review of Genetics, 55*, 633–659. https://doi.org/10.1146/annurev-genet-071719-020506

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., et al. (2017). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution, 9*(2), 410–419. https://doi.org/10.1111/2041-210X.12871

Castañeda-Rico, S., León-Paniagua, L., Edwards, C. W., & Maldonado, J. E. (2020). Ancient DNA from museum specimens and next generation sequencing help resolve the controversial evolutionary history of the critically endangered Puebla deer mouse. *Frontiers in Ecology and Evolution, 8*, 94.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances*, *1*(5), e1400253. https://www.science.org/doi/full/10.1126/sciadv.1400253

Clark, R. D., Catalano, K. A., Fitz, K. S., Garcia, E., Jaynes, K. E., Reid, B. N., et al. (2023). The practice and promise of temporal genomics for measuring evolutionary responses to global change.*Molecular Ecology Resources.* https://doi.org/10.1111/1755-0998.13789

Davis, M., Faurby, S., & Svenning, J. C. (2018). Mammal diversity will take millions of years to recover from the current biodiversity crisis. *Proceedings of the National Academy of Sciences*, *115*(44), 11262–11267. https://www.pnas.org/doi/abs/10.1073/pnas.1804906115

Der Sarkissian, C., Allentoft, M. E., Ávila-Arcos, M. C., Barnett, R., Campos, P. F., Cappellini, E., et al. (2015). Ancient genomics. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences, 370*(1660), 20130387. https://doi.org/10.1098/rstb.2013.0387

Díez-del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., & Dalén, L. (2018). Quantifying temporal genomic erosion in endangered species. *Trends in Ecology & Evolution, 33*(3), 176–185.

Ennes Silva, F., Valsecchi do Amaral, J., Roos, C., Bowler, M., Röhe, F., Sampaio, R., et al. (2022). Molecular phylogeny and systematics of bald uakaris, genus Cacajao (Primates: Pitheciidae), with the description of a new species. *Molecular Phylogenetics and Evolution, 173*, 107509. https://doi.org/10.1016/j.ympev.2022.107509

Estrada, A., Garber, P. A., Rylands, A. B., Roos, C., Fernandez-Duque, E., Di Fiore, A., Nekaris, A-I. et al. (2017). Impending extinction crisis of the world's primates: Why primates matter. *Science Advances*, *3*(1), e1600946. https://www.science.org/doi/full/10.1126/sciadv.1600946

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics, 32*(9), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fa, J. E., Seymour, S., Dupain, J. E. F., Amin, R., Albrechtsen, L., & Macdonald, D. (2006). Getting to grips with the magnitude of exploitation: bushmeat in the Cross-Sanaga rivers region, Nigeria and Cameroon. *Biological Conservation, 129*(4), 497–510. https://doi.org/10.1016/j.biocon.2005.11.03

Fabre, P. H., Rodrigues, A., & Douzery, E. J. P. (2009). Patterns of macroevolution among Primates inferred from a supermatrix of mitochondrial and nuclear DNA. *Molecular Phylogenetics and Evolution, 53*(3), 808–825. https://doi.org/10.1016/j.ympev.2009.08.004

Finstermeier, K., Zinner, D., Brameier, M., Meyer, M., Kreuz, E., Hofreiter, M., & Roos, C. (2013). A mitogenomic phylogeny of living primates. *Plos One, 8*(7), e69504. https://doi.org/10.1371/journal.pone.0069504

Fominka, N. T., Oliveira, H. F. M., Taboue, G. C. T., Luma, F. E., Robinson, C. A., & Fokam, E. B. (2021). Conserving the forgotten: New insights from a Central African biodiversity hotspot on the anthropogenic perception of nocturnal primates (Mammalia: Strepsirrhini). *Primates; Journal of Primatology, 62*(3), 537–546. https://doi.org/10.1007/s10329-021-00898-7

Frantzen, M. A., Silk, J. B., Ferguson, J. W., Wayne, R. K., & Kohn, M. H. (1998). Empirical evaluation of preservation methods for faecal DNA. *Molecular Ecology, 7*(10), 1423–1428. https://doi.org/10.1046/j.1365-294x.1998.00449.x

Freedman, J., van Dorp, L. B., & Brace, S. (2018). Destructive sampling natural science collections: an overview for museum professionals and researchers. *Journal of Natural Science Collections, 5*, 21–34.

Fulton, T. L., & Shapiro, B. (2019). Setting Up an Ancient DNA Laboratory. In B. Shapiro, A. Barlow, P. Heintzman, M. Hofreiter, J. Paijmans, & A. Soares (Eds.), *Ancient DNA Methods in Molecular Biology.* Humana Press. https://doi.org/10.1007/978-1-4939-9176-1_1

Gillespie, T. R., Nunn, C. L., & Leendertz, F. H. (2008). Integrative approaches to the study of primate infectious disease: implications for biodiversity conservation and global health. *American Association of Physical Anthropologists, 137*(S47), 53–69. https://doi.org/10.1002/ajpa.20949

Glander, K. E. (2013). Darting, anesthesia, and handling. In E. Sterling, N. Bynum, & M. Blair (Eds.), *Primate Ecology and Conservation* (pp. 27–39). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199659449.003.0003

Gómez, J. M., Nunn, C. L., & Verdú, M. (2013). Centrality in primate–parasite networks reveals the potential for the transmission of emerging infectious diseases to humans. *Proceedings of the National Academy of Sciences, 110*(19), 7738–7741. https://doi.org/10.1073/pnas.1220716110

Gonçalves, M., Siegismund, H. R., van Vuuren, B. J., Ferrand, N., & Godinho, R. (2021). Evolutionary history of the roan antelope across its African range. *Journal of Biogeography, 48*(11), 2812–2827. https://doi.org/10.1111/jbi.14241

Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J., ..., & Groves, C. P. (1998). Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Molecular Phylogenetics and Evolution*, *9*(3), 585–598. https://doi.org/10.1006/mpev.1998.0495

Groves, C. P. (2001). Primate Taxonomy. Smithsonian series in comparative evolutionary biology. *Smithsonian Institution Press, Washington*, viii, p 350.

Grubb, P., Butynski, T. M., Oates, J. F., Bearder, S. K., Disotell, T. R., Groves, C. P., & Struhsaker, T. T. (2003). Assessment of the diversity of african primates. *International Journal of Primatology, 24*(6), 1301–1357. https://doi.org/10.1023/B:IJOP.0000005994.86792.b9

Guschanski, K., Krause, J., Sawyer, S., Valente, L. M., Bailey, S., Finstermeier, K., et al. (2013). Next-generation museomics disentangles one of the largest primate radiations. *Systematic Biology, 62*(4), 539–554. https://doi.org/10.1093/sysbio/syt018

Hajdinjak, M., Qiaomei, F., Hübner, A., Petr, M., Mafessoni, F., Grote, S., Skoglund, P., et al. (2018). Reconstructing the genetic history of late Neanderthals.". *Nature, 555*(7698), 652–656. https://doi.org/10.1038/nature26151

Hawkins, M. T. R., Bailey, C. A., Brown, A. M., Tinsman, J., Hagenson, R. A., Culligan, R. R., et al. (2022a). Nuclear and mitochondrial phylogenomics of the sifakas reveal cryptic variation in the diademed sifaka. *Genes*, *13*(6). https://doi.org/10.3390/genes13061026

Hawkins, M. T. R., Flores, M. F. C., McGowen, M., & Hinckley, A. (2022b). A comparative analysis of extraction protocol performance on degraded mammalian museum specimens. *Frontiers in Ecology and Evolution*, *10*. https://doi.org/10.3389/fevo.2022.984056

Holmes, I., & Rabosky, A. R. D. (2018). Natural history bycatch: a pipeline for identifying metagenomic sequences in RADseq data. *PeerJ, 6*, e4662. https://doi.org/10.7717/peerj.4662

Jameson, N. M., Hou, Z. C., Sterner, K. N., Weckle, A., Goodman, M., Steiper, M. E., & Wildman, D. E. (2011). Genomic data reject the hypothesis of a prosimian primate clade. *Journal of Human Evolution, 61*(3), 295–305. https://doi.org/10.1016/j.jhevol.2011.04.004

Jensen, A., Swift, F., de Vries, D., Beck, R. M., Kuderna, L. F., Knauf, S., ..., & Guschanski, K. (2023). Complex evolutionary history with extensive ancestral gene flow in an African primate radiation. *Molecular Biology and Evolution*, 40(12), msad247. https://doi.org/10.1093/molbev/msad247

Kebede, T., Bech, N., Allienne, J. F., Olivier, R., Erko, B., & Boissier, J. (2020). Genetic evidence for the role of non-human primates as reservoir hosts for human schistosomiasis. *PLoS Neglected Tropical Diseases*, *14*(9), e0008538. https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0008538

Kistler, L., Ratan, A., Godfrey, L. R., Crowley, B. E., Hughes, C. E., Lei, R., et al. (2015). Comparative and population mitogenomic analyses of Madagascar's extinct, giant "subfossil" lemurs. *Journal of Human Evolution, 79*, 45–54. https://doi.org/10.1016/j.jhevol.2014.06.016

Korlević, P., Gerber, T., Gansauge, M.-T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., & Meyer, M. (2015). Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques, 59*(2), 87–93. https://doi.org/10.2144/000114320

Kotait, I., Oliveira, R. D. N., Carrieri, M. L., Castilho, J. G., Macedo, C. I., Pereira, P. M. C., Boere, V., Montebello, L., & Rupprecht, C. E. (2019). Non-human primates as a reservoir for rabies virus in Brazil. *Zoonoses and Public Health*, *66*(1), 47–59. https://onlinelibrary.wiley.com/doi/abs/10.1111/zph.12527

Kuderna, L. F., Gao, H., Janiak, M. C., Kuhlwilm, M., Orkin, J. D., Bataillon, T., ... & Marques Bonet, T. (2023). A global catalog of whole-genome diversity from 233 primate species. *Science*, *380*(6648), 906–913. https://doi.org/10.1126/science.abn7829

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics, 27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Lo Cantore, P., Giorgio, A., & Iacobellis, N. S. (2015). Bioactivity of volatile organic compounds produced by Pseudomonas tolaasii. *Frontiers in Microbiology, 6*, 1082. https://doi.org/10.3389/fmicb.2015.01082

Lonsinger, R. C., Daniel, D., Adams, J. R., & Waits, L. (2019). Consideration of sample source for establishing reliable genetic microsatellite data from mammalian carnivore specimens held in natural history collections. *Journal of Mammalogy*, *100*(5), 1678–1689. https://www.jstor.org/stable/27009720

Marciniak, S., Mughal, M. R., Godfrey, L. R., Bankoff, R. J., Randrianatoandro, H., Crowley, B. E., ..., & Perry, G. H. (2021). Evolutionary and phylogenetic insights from a nuclear genome sequence of the extinct, giant,"subfossil" koala lemur Megaladapis edwardsi. *Proceedings of the National Academy of Sciences*, *118*(26), e2022117118.

Masters, J. C., Génin, F., Couette, S., Groves, C. P., Nash, S. D., Delpero, M., & Pozzi, L. (2017). A new genus for the eastern dwarf galagos (Primates: Galagidae). *Zoological Journal of the Linnean Society, 181*(1), 229–241. https://doi.org/10.1093/zoolinnean/zlw028

Mauriès, P. (2002). *Cabinets of curiosities*. Thames & Hudson.

McCormack, J. E., Tsai, W. L., & Faircloth, B. C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources, 16*(5), 1189–1203.

McDonough, M. M., Parker, L. D., Rotzel McInerney, N., Campana, M. G., & Maldonado, J. E. (2018). Performance of commonly requested destructive museum samples for mammalian genomic studies. *Journal of Mammalogy, 99*(4), 789–802. https://doi.org/10.1093/jmammal/gyy080

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one, 8*(4), e61217. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0061217

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols, 2010*(6), pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Munds, R. A., Titus, C. L., Eggert, L. S., & Blomquist, G. E. (2018). Using a multi-gene approach to infer the complicated phylogeny and evolutionary history of lorises (Order Primates: Family Lorisidae). *Molecular Phylogenetics and Evolution, 127*, 556–567. https://doi.org/10.1016/j.ympev.2018.05.025

Nekaris, K. A. I., & Burrows, A. M. (Eds.). (2020). Evolution, ecology and conservation of lorises and pottos. Cambridge University Press. https://doi.org/10.1017/9781108676526

Nekaris, K.A.-I., & Nijman, V. (2022). A new genus name for pygmy lorises, Xanthonycticebus gen. nov. (Mammalia, primates). *Zeitschrift für Staats- und Europawissenschaften, 98*(1), 87–92. https://doi.org/10.3897/zse.98.81942

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, ''Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard, D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2022). _vegan: Community Ecology Package_. R package version 2.6-4, <https://CRAN.R-project.org/package=vegan>.

Ozga, A. T., Webster, T. H., Gilby, I. C., Wilson, M. A., Nockerts, R. S., Wilson, M. L., et al. (2021). Urine as a high-quality source of host genomic DNA from wild populations. *Molecular Ecology Resources, 21*(1), 170–182. https://doi.org/10.1111/1755-0998.13260

Parker, L. D., Hawkins, M. T. R., Camacho-Sanchez, M., Campana, M. G., West-Roberts, J. A., Wilbert, T. R., et al. (2020). Little genetic structure in a Bornean endemic small mammal across a steep ecological gradient. *Molecular Ecology, 29*(21), 4074–4090. https://doi.org/10.1111/mec.15626

Parker, L. D., Campana, M. G., Quinta, J. D., Cypher, B., Rivera, I., Fleischer, R. C., et al. (2022). An efficient method for simultaneous species, individual, and sex identification via in-solution single nucleotide polymorphism capture from low-quality scat samples. *Molecular Ecology Resources, 22*(4), 1345–1361. https://doi.org/10.1111/1755-0998.13552

Penna, A., Dillon, R., Bearder, S. K., Karlsson, J., Perkin, A., & Pozzi, L. (2022). Phylogeography and evolutionary lineage diversity in the small-eared greater galago, *Otolemur garnettii* (Primates: Galagidae). *Zoological Journal of the Linnean Society.* https://doi.org/10.1093/zoolinnean/zlac079

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A., ..., & Pecon-Slattery, J. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, *7*(3), e1001342. https://doi.org/10.1371/journal.pgen.1001342

Phukuntsi, M. A., Du Plessis, M., Dalton, D. L., Jansen, R., Cuozzo, F. P., Sauther, M. L., & Kotze, A. (2019). Population genetic structure of the thick-tailed bushbaby (Otolemur crassicaudatus) from the Soutpansberg Mountain range, Northern South Africa, based on four mitochondrial DNA regions. *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis*, 1–10. https://doi.org/10.1080/24701394.2019.1694015

Phukuntsi, M. A, Du Plessis, M., Dalton, D. L., Jansen, R., Sauther, M. L., Cuozzo, F. P., & Kotze, A. (2021). Population and genetic structure of a male-dispersing strepsirrhine, Galago moholi (Primates, Galagidae), from northern South Africa, inferred from mitochondrial DNA. *Primates; Journal of Primatology*. https://doi.org/10.1007/s10329-021-00912-y

Polanc, P., Sindičić, M., Jelenčič, M., Gomerčić, T., Kos, I., & Huber, Đ. (2012). Genotyping success of historical Eurasian lynx (Lynx lynx L) samples. *Molecular Ecology Resources, 12*(2), 293–298.

Porter, L. M., de la Torre, S., Pérez-Peña, P., & Cortés-Ortiz, L. (2021). Taxonomic diversity of Cebuella in the western Amazon: Molecular, morphological and pelage diversity of museum and free-ranging specimens. *American Journal of Physical Anthropology, 175*(1), 251–267. https://doi.org/10.1002/ajpa.24266

Pozzi, L. (2016). The role of forest expansion and contraction in species diversification among galagos (Primates: Galagidae). *Journal of Biogeography, 43*(10), 1930–1941. https://doi.org/10.1111/jbi.12846

Pozzi, L., Disotell, T, R., & Masters, J. C. (2014a) Multilocus phylogeny reveals deep lineages within African galagids (Primates: Galagidae). *BMC Evolutionary Biology 14*, 1–18. https://link.springer.com/article/10.1186/1471-2148-14-72

Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L., & Disotell, T. R. (2014b). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution, 75*, 165–183. https://doi.org/10.1016/j.ympev.2014.02.023

Pozzi, L., Nekaris, K.A.-I., Perkin, A., Bearder, S. K., Pimley, E. R., Schulze, H., et al. (2015). Remarkable ancient divergences amongst neglected lorisiform primates. *Zoological Journal of the Linnean Society, 175*(3), 661–674. https://doi.org/10.1111/zoj.12286

Pozzi, L., Disotell, T. R., Bearder, S. K., Karlsson, J., Perkin, A., & Gamba, M. (2019). Species boundaries within morphologically cryptic galagos: Evidence from acoustic and genetic data. *Folia Primatologica, 90*(5), 279–299.

Pozzi, L., Penna, A., Bearder, S. K., Karlsson, J., Perkin, A., & Disotell, T. R. (2020a). Cryptic diversity and species boundaries within the *Paragalago zanzibaricus* species complex. *Molecular Phylogenetics and Evolution, 150*, 106887. https://doi.org/10.1016/j.ympev.2020.106887

Pozzi, L., Roos, C., & Blair, M. E. (2020b). Molecular advances in lorisid taxonomy and phylogeny. In K. A. I. Nekaris & A. M. Burrows (Eds.), *Evolution Ecology and Conservation of Lorises and Pottos* (pp. 57–66). Cambridge University Press. https://doi.org/10.1017/9781108676526.007

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Accessed 13 Jan 2023. https://www.R-project.org/

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: Advances and challenges in museomics. *Trends in Ecology & Evolution*, *36*(11), 1049–1060. https://www.sciencedirect.com/science/article/pii/S0169534721002147

Rogers, J., & Gibbs, R. A. (2014). Comparative primate genomics: emerging patterns of genome content and dynamics. *Nature Reviews Genetics, 15*(5), 347–359. https://doi.org/10.1038/nrg3707

Rowe, K. C., Singhal, S., Macmanes, M. D., Ayroles, J. F., Morelli, T. L., Rubidge, E. M., ..., & Moritz, C. C. (2011). Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources*, *11*(6), 1082–1092.

Roycroft, E., Moritz, C., Rowe, K. C., Moussalli, A., Eldridge, M. D. B., Portela Miguez, R., et al. (2022). Sequence capture from historical museum specimens: maximizing value for population and phylogenomic studies. *Frontiers in Ecology and Evolution*, *10*. https://doi.org/10.3389/fevo.2022.931644

Rubinoff, D., & Holland, B. S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology, 54*(6), 952–961.

Ruiz-García, M., Sánchez-Castillo, S., Castillo, M. I., Luengas, K., Ortega, J. M., Moreno, P., et al. (2018). How many species, taxa, or lineages of *Cebus albifrons* (Platyrrhini, Primates) inhabit Ecuador? Insights from mitogenomics. *International Journal of Primatology, 39*(6), 1068–1104. https://doi.org/10.1007/s10764-018-0062-6

Schmitt, C. J., Cook, J. A., Zamudio, K. R., & Edwards, S. V. (2019). Museum specimens of terrestrial vertebrates are sensitive indicators of environmental change in the Anthropocene. *Philosophical Transactions of the Royal Society B, 374*(1763), 20170387.

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*, 1–7. https://bmcresnotes.biomedcentral.com/articles/10.1186/s13104-016-1900-2

Shapiro, B., & Hofreiter, M. (2014). A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. *Science, 343*(6169), 1236573. https://doi.org/10.1126/science.1236573

Solórzano-García, B., & Pérez-Ponce de León, G. (2018). Parasites of neotropical primates: a review. *International Journal of Primatology, 39*, 155–182. https://doi.org/10.1007/s10764-018-0031-0

Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., ..., & Murphy, W. J. (2012). Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PloS One*, 7(11), e49521. https://doi.org/10.1371/journal.pone.0049521

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Svensson, M. S., Morcatty, T. Q., Nijman, V., & Shepherd, C. R. (2021). Shedding light on the trade in nocturnal galagos. *Primate Conservation*, *35*(35), 157–168

Thompson, C. W., Phelps, K. L., Allard, M. W., Cook, J. A., Dunnum, J. L., Ferguson, A. W., et al. (2021). Preserve a voucher specimen! the critical need for integrating natural history collections in infectious disease studies. *mBio*, *12*(1). https://doi.org/10.1128/mBio.02698-20

Ting, N., & Sterner, K. N. (2013). Primate molecular phylogenetics in a genomic era. *Molecular Phylogenetics and Evolution, 66*(2), 565–568. https://doi.org/10.1016/j.ympev.2012.08.021

van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. O'Reilly Media.

van der Valk, T., Díez-Del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalén, L. (2019). Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Current Biology, 29*(1), 165-170.e6. https://doi.org/10.1016/j.cub.2018.11.055

Velsko, I. M., Frantz, L. A., Herbig, A., Larson, G., & Warinner, C. (2018). Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research. *Msystems, 3*(4), 10–1128.

Waits, L. P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management, 69*(4), 1419–1433. https://doi.org/10.2193/0022-541X(2005)69[1419:NGSTFW]2.0.CO;2

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the tidyverse. *The Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilkie, D. S., & Carpenter, J. F. (1999). Bushmeat hunting in the Congo Basin: an assessment of impacts and options for mitigation. *Biodiversity & Conservation, 8*, 927–955. https://doi.org/10.1023/A:1008877309871

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology, 20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Woodruff, D. S. (1993). Non-invasive genotyping of primates. *Primates; Journal of Primatology, 34*(3), 333–346. https://doi.org/10.1007/BF02382629

Woods, R., Turvey, S. T., Brace, S., MacPhee, R. D. E., & Barnes, I. (2018). Ancient DNA of the extinct Jamaican monkey Xenothrix reveals extreme insular change within a morphologically conservative radiation. *Proceedings of the National Academy of Sciences of the United States of America, 115*(50), 12769–12774. https://doi.org/10.1073/pnas.1808603115

Yao, L., Witt, K., Li, H., Rice, J., Salinas, N. R., Martin, R. D., et al. (2020). Population genetics of wild Macaca fascicularis with low-coverage shotgun sequencing of museum specimens. *American Journal of Physical Anthropology, 173*(1), 21–33. https://doi.org/10.1002/ajpa.24099

Yoder, A. D. (2013). The lemur revolution starts now: the genomic coming of age for a non-model organism. *Molecular Phylogenetics and Evolution, 66*(2), 442–452. https://doi.org/10.1016/j.ympev.2012.08.024

Yoder, A. D., Cartmill, M., Ruvolo, M., Smith, K., & Vilgalys, R. (1996). Ancient single origin for Malagasy primates. *Proceedings of the National Academy of Sciences, 93*(10), 5122–5126. https://www.pnas.org/doi/abs/10.1073/pnas.93.10.5122

Zacho, C. M., Bager, M. A., Margaryan, A., Gravlund, P., Galatius, A., Rasmussen, A. R., & Allentoft, M. E. (2021). Uncovering the genomic and metagenomic research potential in old ethanol-preserved snakes. *Plos One, 16*(8), e0256353. https://doi.org/10.1371/journal.pone.0256353

Authors and Affiliations

Anna Penna[1,2] · Mary E. Blair[3] · Hsiao-Lei Lui[2] · Elsa Peters[2,4] · Logan Kistler[2] · Luca Pozzi[1]

✉ Anna Penna
pennaa@si.edu

1 Department of Anthropology, University of Texas at San Antonio, San Antonio, TX, USA

2 National Museum of Natural History, Smithsonian Institution, Washington, DC, USA

3 Center for Biodiversity and Conservation, American Museum of Natural History, New York, NY, USA

4 Licenciatura en Ciencias Genómicas, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México Campus Morelos, Cuernavaca, Morelos, Mexico