Combining eDNA and Museomics to Enhance Biodiversity Monitoring

Sarah Schmid[1,2], Nicolas Straube[3], Camille Albouy[1,2], Bo Delling[4], James Maclaine[5], Michael Matschiner[6,7,8], Peter Rask Møller[9], Annamaria Nocita[10], Anja Palandačić[11,12], Lukas Rüber[13,14], Moritz Sonnewald[15], Nadir Alvarez[16,17], Stéphanie Manel[18,19], Loïc Pellissier[1,2]

[1]Ecosystems and Landscape Evolution, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland

[2]Land Change Science Research Unit, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

[3]University Museum Bergen, University of Bergen, Bergen, Norway

[4]Department of Zoology, Swedish Museum of Natural History, Stockholm, Sweden

[5]Natural History Museum, London, United Kingdom

[6]Natural History Museum, University of Oslo, Oslo, Norway

[7]Ludwig Maximilian University of Munich, Munich, Germany

[8]Bavarian State Collection for Zoology, Munich, Germany

[9]Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

[10]Natural History Museum, Florence University, Florence, Italy

[11]Natural History Museum Vienna, Vienna, Austria

[12]Department of Biology, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

[13]Naturhistorisches Museum Bern, Bern, Switzerland

[14]Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

[15]Senckenberg Research Institute, Department of Marine Zoology, Section Ichthyology, Frankfurt am Main, Germany

[16]Naturéum – State Museum of Natural History, Palais de Rumine, Lausanne, Switzerland

[17]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

[18]CEFE, Univ Montpellier, CNRS, EPHE-PSL University, IRD, Montpellier, France.

[19]Institut Universitaire de France, Paris, France

Corresponding author: Sarah Schmid (sarah.schmid@ik.me)

Abstract

Biodiversity changes due to human activities highlight the need for efficient biodiversity monitoring approaches. Environmental DNA (eDNA) metabarcoding offers a non-invasive method to assess species distributions, but its accuracy depends on comprehensive DNA reference databases. Natural history museum collections often contain rare or difficult to obtain samples that could be used as a resource to fill gaps in eDNA reference databases. In the present paper, we evaluated the potential of retrieving DNA sequences from fish museum specimens using commercial kits and custom protocol for museum specimens. We then discuss how museomics – the application of -omics techniques to museum specimens – not only has the potential to improve eDNA reference databases but could also lead to transferable methodological advancements. Combining the field of museomics and eDNA could enhance our understanding of global biodiversity and highlight the value of natural history collections.

Keywords

Natural history collections, fish, metabarcoding, DNA reference database, historical DNA

Challenges in biodiversity monitoring: the potential of eDNA metabarcoding and the importance of reference database completeness

We are witnessing a significant change in Earth's biological diversity driven by anthropogenic factors (Pereira et al. 2010), resulting in geographic redistribution of species (Lenoir et al. 2020, Chevalier et al. 2024) and spatiotemporal re-organisations of communities in both terrestrial and aquatic biomes (Walsh et al. 2015, García-Navas et al. 2020). Considering the speed of the environmental modifications induced by global changes, it is crucial to assess the shift in the distribution of various taxonomic groups and pinpoint areas where species are most at risk to enable effective planning and resource allocation. Traditional monitoring survey methods often miss elusive or rare species (Boussarie et al. 2018, Mathon et al. 2022), suffer from a geographic bias in sampling efforts, and are not well-suited to detect rapid modifications of species community composition induced by climate change (Staudinger et al. 2013). Environmental DNA (eDNA) metabarcoding has been proposed to gather present-day species occurrences for various taxa and ecosystems more efficiently (Valentini et al. 2016, Pereira et al. 2021) and facilitate the monitoring of anthropogenic impact on biodiversity.

eDNA metabarcoding has the potential to speed up the collection of species distribution information, as it involves analysing genetic material obtained from environmental samples (eDNA) that contain a mixture of intra- and extracellular DNA, without the need to collect individuals from the ecosystem (Taberlet et al. 2012). The process involves species detection through water or air filtration (Clare et al. 2022), soil sampling (Allen et al. 2023), or surface swabbing (Aucone et al. 2023), followed by amplifying and sequencing of one or more DNA barcodes (see below), which are then compared to a genetic reference database for species identification (Fraija-Fernández et al. 2020). This non-invasive approach has been applied to detect a range of organisms, including fishes (Ramírez-Amaro et al. 2022, Rozanski et al. 2022), and has demonstrated the ability to detect species occurrences of elusive species like sharks (Bakker et al. 2017, Boussarie et al. 2018), large pelagic species (Veron et al. 2023) and cryptic species such as gobies (Boulanger et al. 2021). To make eDNA metabarcoding a truly effective tool for conservation, it is essential to accurately assign eDNA sequences to the correct taxa, a process that depends on comprehensive DNA reference databases (Keck et al. 2022).

DNA reference databases consist of short DNA sequences (usually between 100 to 700 base pairs) that are taxonomically annotated and curated. If incomplete, a reference database will

jeopardise the integrity of eDNA-based biodiversity assessment in a given ecosystem, especially if it misses important species for conservation. Despite substantial efforts to increase the number of barcode and genomic sequences available (figure 1), existing genetic reference databases remain incomplete (Marques et al. 2021). This gap impedes our ability to assign eDNA sequences to the correct species and thus remains an obstacle to the use of eDNA as a tool for biodiversity monitoring (Beng and Corlett 2020). Scientific consortia and working groups are currently working globally to create and maintain high-quality reference databases (e.g., MIDORI2, PFR2, dinoref, PhytoREF, Mare-MAGE). In addition, an increasing number of projects aims at sequencing the entire genomes of specific target species sets (e.g., Global Invertebrate Genomics Alliance, Earth BioGenome Project, European Reference Genome Atlas, Vertebrate Genomes Project; see (Formenti et al. 2022) for references). These programs seek to offer novel perspectives on genomic diversity and architecture, but they also substantially advance the field of eDNA research by supplying the entire spectrum of barcodes necessary for species identification in metabarcoding analysis. Although samples for such projects can be obtained from biobanks, collection of individuals *in situ* is more frequently used, which is time-consuming and requires taxonomic expertise to achieve accurate species identification. Furthermore, because missing species in databases are often rare and elusive, sampling living organisms presents not only practical challenges and financial burdens but also ethical issues. An alternative would be to leverage the wealth of scientific collection material preserved in natural history museums – particularly type specimens, the individual specimens species descriptions are based on, or other rigorously identified specimens (Renner et al. 2024) – to improve eDNA reference databases.

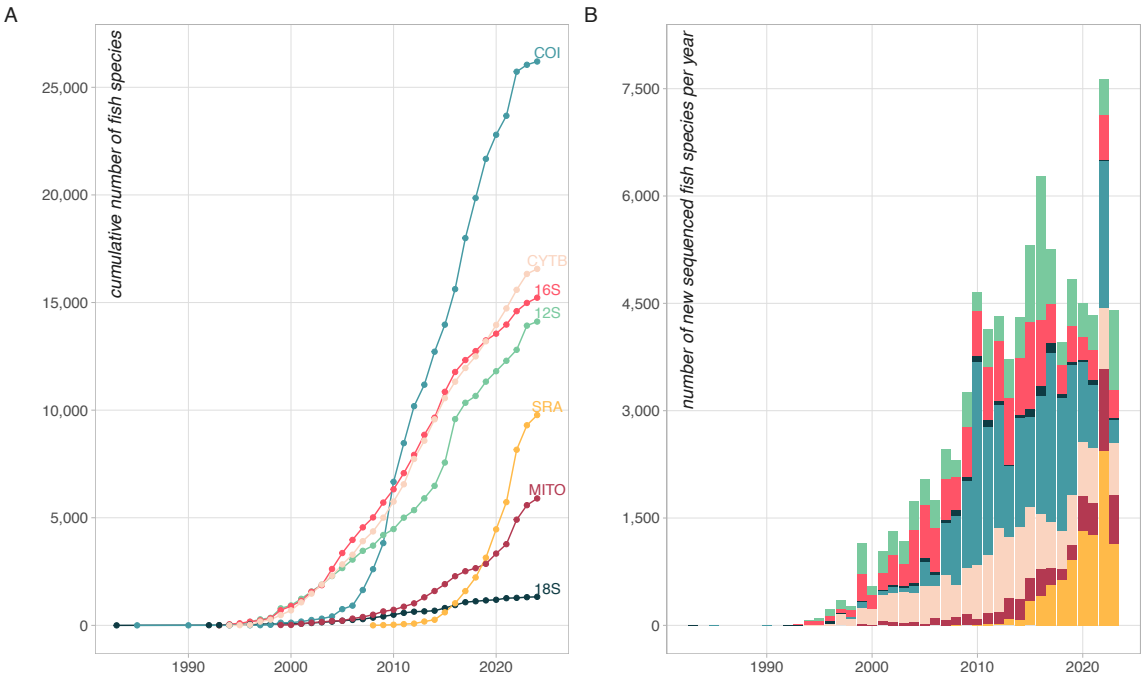

Figure 1. Cumulative number of species sequenced for a given barcodes/type of genomic data. (A) The lines correspond to the cumulative number of fish species with a given barcode or genomic data available. The values were retrieved from NCBI for all fish species (marine and freshwater). (B) The bar plots correspond to the number of new species added each year for a given barcode or genomic data. COI: cytochrome oxidase I (blue), CYTB: cytochrome B (light pink), 16S: 16S ribosomal RNA (pink), 12S: 12S ribosomal RNA (light green), 18S: 18S ribosomal RNA (dark blue), MITO: complete mitochondrial genome (dark red), SRA: sequence read archive, which consists in various types of genomic data (yellow).

Natural history museum collection and museomics in support of environmental problems

The use of taxonomically identified specimens from natural history museum collections could speed up the establishment of reference databases if we overcome the technical challenges associated with the analysis of those specimens. Museum collections can be divided into two categories: dry and wet collections. Dry collections consist of samples that have been dried rather than being stored in a liquid preservation solution such as formalin or ethanol. These collections can include specimens such as pinned insects, stuffed fish, bones or pressed plants. Wet collections, in contrast, consist of specimens which might have been fixed with formaldehyde and subsequently stored in a preservative liquid such as 75% ethanol. Wet specimens can include a variety of organisms, from small invertebrates to larger animals like fish or reptiles, as well as organ or tissue samples. Such collections are therefore exceptional repositories of taxonomic knowledge (Winker 2004), genetic source material (Wandeler et al. 2007) and historical and ecological data (Fong et al. 2023, Jones et al. 2024). They can thus enhance our understanding of species distribution over time (Elith et al. 2006, Baer et al. 2023) or morphological adaptation to climate change (MacLean et al. 2018), and they can inform the assessment of a species' conservation status (Mollen and Iglésias 2023). Their contributions extend beyond natural sciences research, in fields such as public health (Cook et al. 2020) or education (Ellwood et al. 2020, Leerhøi et al. 2024). The long-term value of these collections to society emphasises the need for their preservation over time, since they may hold unforeseen benefits (Miller et al. 2020) even as their funding decreases (Bradley et al. 2014). Indeed, numerous studies are built on samples that were initially gathered for a different purpose. With technical advancements, they can reach new audiences and provide answers to novel scientific questions (Meineke et al. 2018, Heberling et al. 2019, Lauridsen et al. 2022, Davis and Knapp 2024). This hidden potential has led to the emergence of the term "Museomics" – the application of -omics techniques (e.g., genomics, epigenomics) to museum specimens – around the year 2009, even though museum specimens have been collected for

at least three centuries. Its application resulted in more comprehensive phylogenomic studies (Ruane and Austin 2017, McGuire et al. 2018, Wood et al. 2018, Lyra et al. 2020), tracking the genetic response of species to recent environmental changes (Bi et al. 2019, Byerly et al. 2022), solving taxonomic uncertainties (Muschick et al. 2022, Renner et al. 2024) and more generally reconstructing the evolutionary processes, from population-level to macroevolutionary analyses (Bi et al. 2013, Burrell et al. 2015).

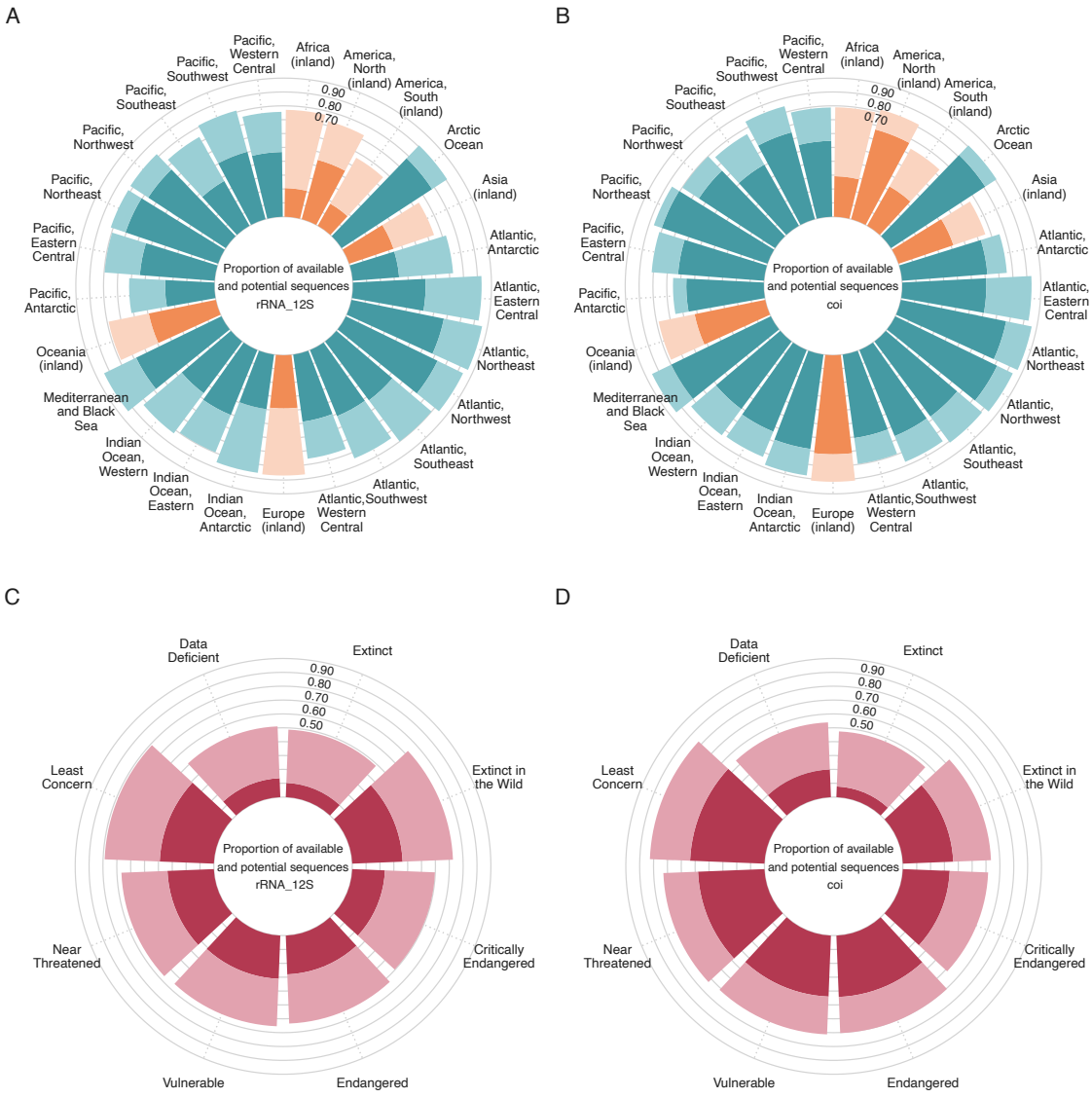

Figure 2. Potential of fish museum specimens to improve reference database for two highly used mitochondrial genes in eDNA studies (12S ribosomal RNA (12S) and cytochrome C oxidase I (COI); other genes are displayed in supplemental material figures S1 and S2). Proportion of sequences available is highlighted in dark colour. Potential new sequences based on available museum specimens in European collections are displayed in lighter colour. (A) Potential of improvement for 12S gene according to geographic regions. Regions in blue are hosting marine fish species while regions in orange

are for freshwater species (inland regions). (B) Potential of improvement for COI gene according to geographic regions. Regions in blue are hosting marine species while regions in orange are for freshwater species (inland regions). (C) Potential of improvement for 12S gene according to global IUCN Red List of Threatened Species assessments for marine and freshwater fishes. (D) Potential of improvement for COI gene according to global IUCN Red List of Threatened Species assessments for marine and freshwater fishes. The complete list of fish species was retrieved from FishBase using the R package rfishbase (Boettiger et al. 2012) as well as their corresponding geographic distribution. The museum data were retrieved from GBIF. The IUCN Red List assessments were retrieved from the IUCN website.

Collections at natural history museums also hold the promise of enhancing reference DNA sequence databases for species worldwide, including threatened ones (figure 2, figures S1-S2). Fresh DNA tissues stored in natural history museum collections (i.e. voucher specimens) have demonstrated significant potential to augment eDNA reference databases (de Santana et al. 2021, Bemis et al. 2023). However, leveraging historical specimens to enhance these databases is a recent advancement (Levesque-Beaudin et al. 2023). Although historical museum specimens were not originally collected for molecular research – making it challenging to retrieve genomic data from them (see box 2) – these specimens can serve as the only available source of genetic information for rare, elusive, or extinct species that are difficult to sample in the field. Obtaining genomic data from type specimens would be the ideal approach, as DNA sequences derived from these authoritative references are inherently named correctly, even as taxonomic perspectives shift and necessitate transferring names to different genera or altering the rank of taxa (Renner et al. 2024). Furthermore, museum specimens serve as authoritative vouchers, enabling the direct linkage of DNA sequences to the original specimen. This ensures reproducibility in genetic studies and prevents the propagation of taxonomic errors (Buckner et al. 2021). Finally, natural history museum collections frequently contain multiple specimens of a given species sampled across its geographic distribution. This diversity of specimens is valuable for eDNA databases, as it allows capturing the intraspecific genetic variability of species. By incorporating multiple sequences per taxon (e.g. from different geographic locations), the eDNA reference database can be enhanced to more comprehensively represent the target species, ensuring more accurate species identification (Blackman et al. 2023).

Box 2. Addressing the challenges of using museum specimens in molecular biology research

Modern natural history museum collections typically include both historic material and recent tissues, which have the potential to enhance genetic reference databases for eDNA

metabarcoding. The vast majority of these collections consist of historic specimens, whose curatorial treatment was not oriented towards preserving DNA and frequently involves formaldehyde fixation, particularly for fish specimens (Simmons 2014). Formaldehyde treatments are efficient in preserving morphological characters, but they damage DNA, inducing cross-linking, hydrolysis and methylol adducts, which inhibit multiple molecular techniques required to access the DNA sequence (e.g., (Ruane and Austin 2017). Furthermore, the handling history of such museum specimens is frequently undocumented, leading to difficulties to ascertain the techniques used for preserving them. For instance, it is often unclear whether the specimens were preserved in ethanol or formaldehyde (or sometimes both), and for what duration, leading to significant variation of the degree of DNA preservation across different specimens (Cook et al. 2014, McDonough et al. 2018). In consequence, wet collection specimens – as opposed to dry collection specimens for which high sequencing success rates are usually achieved – remain challenging to process, and optimised protocols are required for the extraction and subsequent library preparation of the historical DNA (Raxworthy and Smith 2021).

DNA extraction

The formaldehyde fixation problem was initially addressed in the medical field, where histological cuts are routinely fixed with formaldehyde and commercial kits were developed to tackle the cross-linking issue (e.g., QIAamp DNA FFPE Tissue Kit, QIAGEN; EchoLUTION FFPE DNA Kit, BioEcho). However, museum specimens may have been fixed with higher concentrations of formaldehyde and over extended periods, leading to further degradation of DNA. As a result, specific protocols have been tailored for museum specimens, usually involving the recovery of small DNA fragments and the disruption of formaldehyde-induced cross-links (Straube et al. 2021, Hahn et al. 2022). Such custom protocols are often more time-consuming and require a high financial investment to be established in a lab. The possibility to use commercial kits can thus be appealing, especially when not planning on working in the long-term with museum specimens, but rather for a limited period or number of specimens (e.g., for a single project). To our knowledge, no commercial DNA extraction kit exists specifically for museum specimens, but kits for formalin-fixed paraffin-embedded (FFPE) tissues developed in the medical field can be an alternative.

DNA library preparation

In addition to DNA extraction, library preparation for museum specimens also requires tailored protocols. The fragmented and degraded nature of DNA from museum specimens necessitates modifications to standard library preparation methods. Library preparation for degraded DNA has been widely explored in the field of ancient DNA (Gansauge et al. 2020, Kapp et al. 2021). An important methodological advancement involves protocols for creating single-stranded DNA (ssDNA) instead of double-stranded DNA libraries (dsDNA) (Wales et al. 2015). This technique includes damaged molecules, such as those with abasic sites, and short molecules that are often lost during dsDNA library preparation. While preparing ssDNA libraries is more time-consuming and costly compared to dsDNA protocols, these have played an essential role in several ancient DNA and historical DNA studies (Dabney et al. 2013, Gansauge and Meyer 2013). Such DNA library preparation protocols have inspired recently developed protocols dedicated to museum samples (Straube et al. 2021) and have now been made available in the form of commercial kits (xGen ssDNA & Low-Input DNA Library Prep Kit, IDT; SRSLY NGS Library Prep Kit, ClaretBio). The DNA libraries can then either be directly sequenced or enriched for the taxon of interest based on hybrid-capture protocols, which target homologous DNA regions at broad phylogenetic scales based on a single set of probes (Lemmon and Lemmon 2013, Agne et al. 2022).

Optimising DNA sequence recovery from museum specimens to expand eDNA reference databases

Museum specimens have the potential to provide DNA sequences for a wide range of species, including those that are rare, endangered, or presumed extinct, as well as species from regions under-represented in sequenced barcodes (figures 2, S1-S2). To further illustrate this potential, we extracted DNA, constructed DNA libraries, and sequenced the genome of ten marine fish specimens from collections across Europe, encompassing a diverse range of taxa including bony fishes, sharks, rays, and chimaeras, many of which are elusive or threatened species (table S2). We specifically aimed to evaluate the effectiveness of commercial kits for DNA extraction and library preparation in comparison to a customised museomics protocol (Straube et al. 2021; see supplemental material and methods for more information). These wet-collection specimens were gathered between 1852 and 1993 and their preservation history is mostly unknown. On the other hand, museum specimens collected after the early 20th century were likely formalin-fixed prior to long-term storage in ethanol, as this method became common practice in natural history collections during that time period (see box 1).

Box 1. Obtaining tissue samples from natural history museum collections for molecular analyses

Natural history museums increasingly welcome opportunities to use their collections for scientific research purposes. However, sampling tissues from these collections for molecular analyses necessitates careful consideration, both to prevent contamination of the sampled tissue and to maintain the integrity of the collection. Here, we offer guidance on the protocol to be followed when obtaining fish tissues from wet collections. The following recommendations serve as suggestions and should be tailored to the specific project objectives, always in consultation with the collection curator.

Tools and preparation of the sampling space

The potential presence of exogenous DNA from bacteria, fungi, or due to human handling is a common concern when working with museum specimens. To limit further human-introduced contaminants and cross-contamination between samples, it is critical to collect tissue samples in the most sterile manner possible. Prior to sampling, the workspace should be thoroughly cleaned using DNA removal solution or bleach. Additionally, all tools used must be cleaned between samples (e.g., bleach, flame sterilisation), and gloves should be worn and changed regularly throughout the procedure. Standard dissection equipment, as well as UV-treated 1.5/2 ml tubes, are necessary for this procedure. Careful consideration is needed when using bleach and ethanol simultaneously, as their combination can generate chloroform.

Identifying suitable specimens for sampling

Fish specimens from wet collections often undergo formalin fixation, which can significantly compromise DNA integrity and impair DNA extraction and subsequent library preparation (see box 2). Although adapted protocols are available to retrieve DNA from such specimens (Straube et al. 2021, Hahn et al. 2022), it is recommended to prioritise specimens preserved with alcohol-based fixative to enhance the chances of obtaining genomic data (Duval et al. 2010). Although we often lack information on the preservation history of specimens, we can still rely on other clues to assess the potential DNA sequencing success of a specimen. Based on the specimen characteristics, bleached eye lenses resulting in almost white coloration can indicate pure ethanol preservation whereas dark eye lenses often indicate formalin-fixation prior to ethanol preservation of the specimen, preventing bleaching of the eyes (De Bruyn

et al. 2011). The collection date of the specimen can also inform about the fixative treatment, since formalin was not used in natural history museums until after 1900 (Simmons 2014), meaning that specimens collected before the 20th century were probably not fixed with formalin. Furthermore, residual formaldehyde concentration in the storage solution can be quantified in addition to its pH, as lower pH is expected when specimens are formalin-fixed, due to the breakdown of formaldehyde into formic acid (Hahn et al. 2022). Although these indicators can provide clues about preservation, their lack cannot guarantee the absence of formalin fixation. When feasible, it is thus advisable to sample multiple specimens of the same species, which come from different lots (with different collection year, collector and so on), to increase the likelihood of successful DNA sequencing.

Selecting the appropriate tissue sample

The ideal amount of tissue used for DNA extraction may vary from sample to sample and is also limited by the amount of tissue that can be sampled without damaging the specimen. An excessive amount of tissue can block the DNA extraction column and lead to a higher concentration of potential DNA extraction/PCR inhibitors. Even though we generally lack experimental data correlating the amount of tissue to the extracted DNA concentration and quality, a safe option would be to extract several small-sized tissue samples from the same specimen in parallel and combine them afterwards. For fish specimens, the optimal sampling location is the right flank, as the left flank is used for morphometric analysis. Often, specimens exhibit a cut on the belly (for formalin-fixation purposes), allowing for the sampling of internal organs or muscle tissue. Internal organs such as liver tissue have been shown to yield more DNA compared to muscle and skin (Hahn et al. 2022, Palandačić et al. 2024). If the fish specimen remains completely intact, gill filaments can be utilised. Alternatively, a small hole can be created in the skin using a biopsy punch, or a thin incision can be made to retrieve tissue from beneath the skin using forceps. For particularly small or precious specimens, it is possible to proceed to needlepoint non-destructive internal tissue sampling as described in (Haÿ et al. 2020).

Recommendations for tissue storage

Following specimen collection, the tissue should be stored in 95% ethanol. Ethanol with a lower concentration contains more water, which may negatively impact tissue preservation. To mitigate ethanol evaporation, the utilisation of screw-top tubes with a rubber seal is recommended, in addition to long-term storage at -80°C when possible (the colder the better). Further evaporation prevention can be achieved by sealing the tube with parafilm.

For this pilot laboratory investigation, we retrieved low-coverage genomic data from the ten fish species using both protocols and compared their performances by mapping the reads to the closest reference genome available. Sequencing DNA from wet-preserved fish specimens appears achievable, and both commercial and customised protocols can be employed for this purpose. With only ten samples, our findings provide descriptive statistics and basic insights rather than definitive conclusions, and a larger study would be necessary to validate these observations. Nonetheless, we did not observe a difference in the number of reads mapped to the reference genome when comparing the commercial and custom methodologies (figure 3). The decision to use a custom protocol or a commercial kit for museum specimens depends on various factors. Using commercial kits requires less expertise in the molecular lab as most reagents are ready to use, unlike custom protocols which involve preparing solutions for each step, making these time-consuming, potentially prone to errors and increasing the chances of sample contamination. On the other hand, although there is an initial financial investment in purchasing stock reagents for custom protocols, the material costs per sample are usually lower compared to commercial kits (respectively €42.90 and €31, not including quantification and sequencing costs). Costs for working hours, however, may be increased due to the differences in hand-on times. Therefore, choosing between the two methods depends on whether the lab procedure will be a long-term implementation or just a one-time project. For punctual projects with limited samples, commercial kits can be an advantageous approach due to their ease of use and speed, whereas custom protocols might be better suited for long-term projects not only financially but also because they can easily be adjusted according to the project needs.

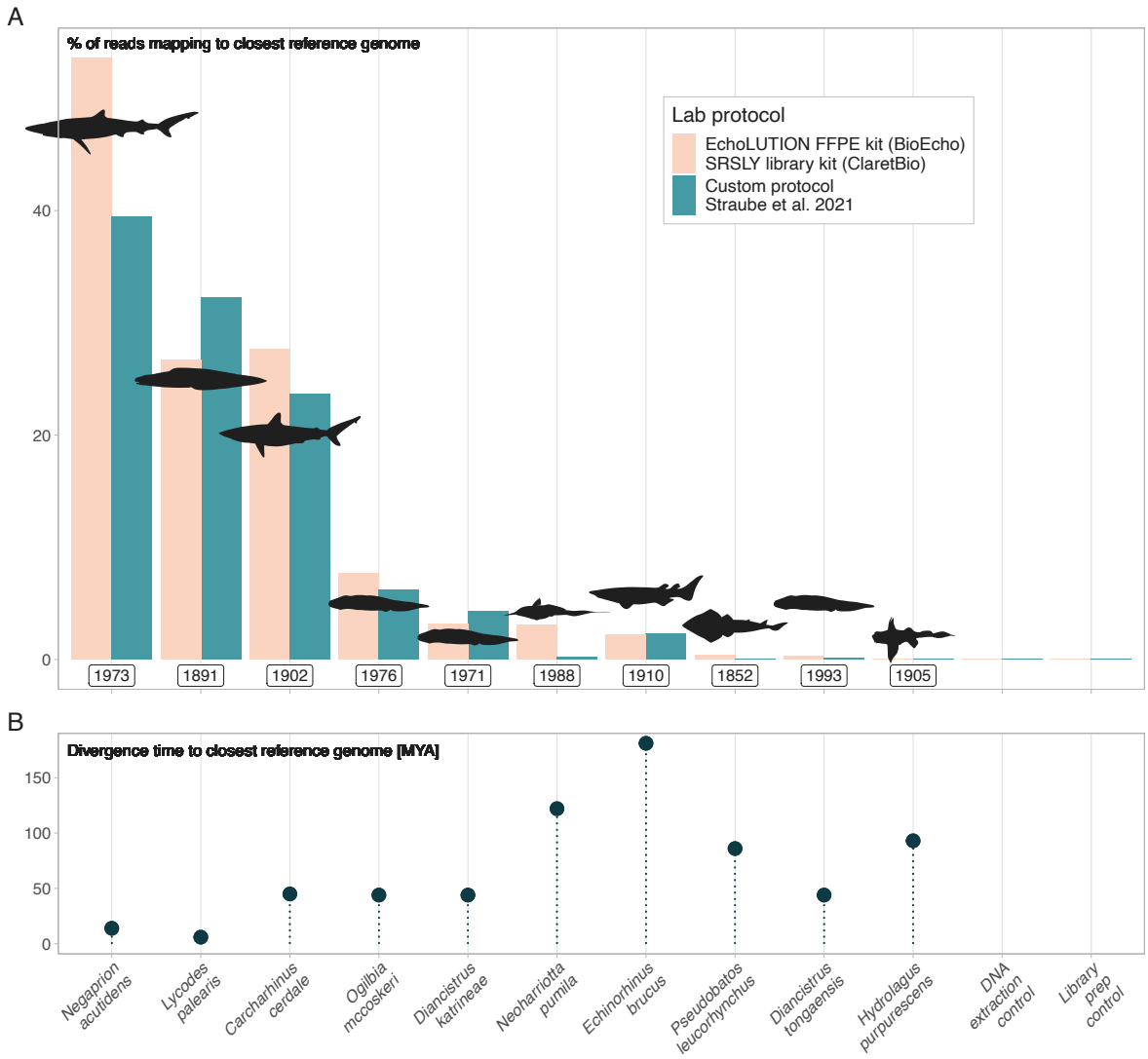

Figure 3. (A) Percentage of reads mapping to the closest reference genome available (divergence between target species and reference genome ranging from 6 to 181 million years ago (MYA); see table S2). Light pink bars highlight results based on the "commercial kit" protocol and blue bars are for the "custom" protocol. The value below each pair of bar plots correspond to the collection date of the museum specimens. (B) Divergence time to the closest reference genome for each species.

While the custom and commercial protocols showed comparable performance, we observed significant variation in the number of mapped reads to reference genomes across samples (figure 3). This discrepancy may be attributed to mapping challenges, such as the selection of an unsuitable reference genome. The divergence time between the samples and their closest reference genome varied substantially (from 6 to 181 million years ago, estimated with TimeTree (Kumar et al. 2022)), and this factor appears to have a considerable impact on the percentage of reads mapping to the reference (figure S3). When generating new genomic data for species lacking a reference genome and working with degraded DNA, the availability of reference genomes is an important consideration. The fragmented nature of museum

specimen DNA may indeed complicate *de novo* reconstruction of the nuclear genome. However, when the goal is enhancing eDNA reference databases, a complete mitochondrial genome (also referred as mitogenome) sequence may be sufficient and more readily reconstructed. Indeed, computational tools like MITObim (Hahn et al. 2013) can be valuable in reconstructing sequences, even without a closely related full reference genome, as they can facilitate the reconstruction of full mitochondrial genomes starting from a single barcode sequence.

The divergence from the reference genome may only partially account for the limited number of reads mapping to the reference for certain species. The low mapping success could also be linked to compromised DNA integrity in the samples, resulting in the failure to generate DNA libraries of sufficient quality for sequencing in some cases. Such sequencing failures may indicate severe DNA damage, potentially related to specimen handling, as well as a high proportion of exogenous DNA being sequenced. Contamination is a significant concern with museum specimens, as historical DNA can be contaminated with DNA from various exogenous sources (e.g., due to human handling, bacteria, or fungi) or from recycling preservation fluid between specimens. Although recent contamination can be minimised during tissue sampling (see box 1), past contamination might dominate over endogenous DNA during the sequencing. While bioinformatic decontamination tools, such as Kraken (Wood and Salzberg 2014), can help identify and remove contamination after sequencing, preventing contamination from being sequenced will yield higher quality data. Laboratory procedures like hybrid-capture are a solution to enrich endogenous DNA and reduce the amount of exogenous DNA sequenced (Gasc et al. 2016).

Capture protocols involve creating a molecular probe — a DNA sequence complementary to the target region of interest — to capture desired genomic regions while minimising exogenous DNA contamination. Various hybrid-capture protocols have been developed for targeting specific regions of interest in the genome while minimising contamination from exogenous DNA. For example, fully customised probes can be manufactured by synthesis platforms and various commercial kits now exist to allow in-solution hybridisation capture. The probes can also be generated from fresh specimens using different methods, such as long-range PCR (Bekaert et al. 2016, González Fortes and Paijmans 2019), RAD-seq derived loci (Suchan et al. 2016) or exon amplification (Schmid et al. 2017). By selectively capturing and amplifying specific genomic targets, these approaches reduce contamination impact and enhance the efficiency of sequencing DNA from museum samples. The probes required for this amplification are still expensive to produce, but their costs might decrease in the future, notably thanks to the invention of benchtop DNA printers (Grinstein 2023).

Potential of museum specimens: beyond improving eDNA reference database

Similar challenges leading to transferable methodological advancements

Museum collections offer promising opportunities to enhance eDNA reference databases. Yet, the use of museum specimens in eDNA research extends beyond augmenting the number of barcodes available. Indeed, some parallels can be drawn between the eDNA and the museomics fields. As mentioned, DNA from archival museum specimens is degraded and potentially contaminated by various sources of exogenous DNA. Such characteristics are shared with eDNA samples, which also display various stages of degradation. Therefore, techniques developed for extracting and sequencing the often-degraded DNA from museum specimens can be adapted for use in eDNA studies. Additionally, understanding the degradation processes of DNA in museum samples can inform about similar challenges faced in eDNA studies, such as variances in degradation due to environmental factors, or even deamination patterns. eDNA samples consist of a mixture of DNA from different organisms, including target species, non-target species, and exogenous contaminants. Therefore, applying capture methods used for sequencing museum specimens (see above) to eDNA studies has the potential to help improve the accuracy and reliability of eDNA-based biodiversity assessments. A capture-based method could also potentially result in a PCR-free eDNA approach, resolving the issues associated with PCR amplification biases (Piñol et al. 2015).

The field of eDNA capture is relatively new and holds great potential for improving the efficiency and accuracy of metabarcoding studies. Unlike most museomics research, eDNA studies target a wide spectrum of species that may exhibit high divergence. This raises the question of which DNA probe sequences to use. One strategy involves utilising barcode gene DNA from a distantly related species that is approximately equidistant to all the studied species (Mariac et al. 2018). Another approach involves aggregating available DNA sequences from databases and employing *in silico* approaches to generate a set of probes that summarises the information within those sequences. This method has been applied at both barcode (Lentz et al. 2021, Agne et al. 2022, Günther et al. 2022) and mitogenome levels (Seeber et al. 2019, Li et al. 2023), as well as for ultraconserved elements (UCEs, Geburzi et al. 2024) and universal single-copy orthologs (USCOs, Dietz et al. 2022).

Shifting away from the single-barcode approach

A key question is which genetic marker will be most suitable for future eDNA studies of fish. Previous research has indicated that the COI gene is suboptimal, due to the lack of primer sites that can be used for species-specific amplification. The 12S gene has been shown to serve as a more suitable alternative, but the associated reference databases are less complete (Collins et al. 2019). Also, the use of a single mitochondrial barcode makes it challenging to reliably distinguish closely related species. Alternatively, eDNA analysis across multiple gene regions could facilitate detailed ecological assessments and identification of indicator species (Seymour et al. 2020). This multilocus approach has demonstrated the capacity to capture a wide range of organisms and yields reliable taxonomic information (Andres et al. 2021) and the ability to obtain population-level genetic data (Andres et al. 2023). The transition from single barcode to a multilocus approach in eDNA metabarcoding is further facilitated by the development of bioinformatic tools enabling the efficient processing of multilocus metabarcode data and the generation of comprehensive reference databases (Curd et al. 2019), as well as the availability of multilocus universal primer sets (Wang et al. 2023). While the specific genetic markers used in future eDNA metabarcoding may vary, sequencing complete mitochondrial genomes could represent a more efficient strategy to enable effective metabarcoding in the long term (but see also (Funk and Omland 2003).

Full mitochondrial genomic information can be a valuable resource for reliable species identification (Dziedzic et al. 2023) and for developing primers for single-species and metabarcoding assays. Whole-mitogenome data are also necessary for transitioning from metabarcoding/barcoding to capture enrichment (Wilcox et al. 2018) or PCR-free environmental genomics. Such methods are not only addressing issues related to PCR amplification biases (Piñol et al. 2015)) but also enable precise quantification of relative species abundance in samples (Yang et al. 2021). Reference databases including full mitogenomes will also enable the analysis of long-read sequencing output. Long-read sequencing starts to be associated with barcoding for species identification (e.g., haplotagging) under fieldwork conditions (Krehenwinkel et al. 2019). Using long-read sequencing could enable the precise capture of taxonomic and phylogenetic diversity within biological communities. Therefore, mitochondrial sequencing and analysis can improve the accuracy of species identification and enable more efficient monitoring of biodiversity using eDNA.

Building comprehensive and carefully curated databases

The availability of well-curated reference databases which are error-free and taxonomically verified is still an issue in the field of eDNA (Keck et al. 2022). Public databases such as

GenBank can be helpful (Leray et al. 2019), but sequence data lack uniform linkage to taxonomically verified vouchers, weakening the connection between DNA sequence and taxonomic identity (Locatelli et al. 2020). While quality-checking at GenBank has improved, the sequence data it holds can still be of draft quality and may contain errors, particularly at the species or subspecies level (Meiklejohn et al. 2019). RefSeq, on the other hand, provides curated and well-annotated sequence datasets but is not comprehensive for all species. Alternatively, the Barcode of Life Data System (BOLD) has more stringent voucher criteria and consistently checks for contaminants, low-quality records, pseudogenes or data irregularities (Ratnasingham and Hebert 2007). However, BOLD relies heavily on information from the cytochrome c oxidase I sequence (COI), which may have limitations in discerning recently diverged sister species pairs and amplifying certain taxa due to poorly conserved primer binding regions (Deagle et al. 2014).

To enhance the potential of eDNA for biodiversity monitoring, a promising approach could be to shift the focus towards the development of extensive and curated mitogenomic databases. With repositories containing vouchered samples and comprehensive mitogenomic data, researchers can access the necessary genetic information for the effective use of eDNA in biodiversity studies (de Santana et al. 2021). Efforts to improve mitogenome databases are ongoing (e.g., Zhu et al. 2023), and guidelines are available to create regional databases of mitogenome sequences for target taxa (Dziedzic et al. 2023). Consistent with the ongoing efforts to sequence new mitogenomes, utilising museum specimens for this purpose could be a complementary approach, as it would facilitate the sequencing of species that are challenging to collect in the field. Additionally, metadata related to the specimen as well as morphological data could be linked to the mitogenome sequence in the database, which is a step towards scientific collections digitisation as well as towards the “extended specimens concept” (2017, Chang 2020). An extended specimen includes not only the physical specimen itself, but also associated data such as measurements, environmental information, photographs/x-rays, DNA sequences, and parasites/symbionts found on the material. By explicitly linking the specimen to other data – including DNA sequences and subsequent analysis incorporating the specimen's information – it could help define the level of confidence we can have in a given sequence used for eDNA-based research. Knowing that a barcode sequence originated from a type specimen would significantly increase confidence in that sequence and the subsequent analyses derived from it, going beyond the capabilities of current eDNA databases (i.e., accuracy of species-level identifications and cryptic lineages is significantly increased).

Considerations regarding the use of museum specimens

Although museum specimens can serve as valuable resources for enhancing eDNA reference databases, certain limitations exist. One of the key limitations is related to reliable taxonomic identification. When working with historical collections, the taxonomic classifications can be inaccurate, as species definitions and concepts have evolved, and species identification provided by the museum may not always be done by expert taxonomists. To overcome sequencing misidentified specimens, the optimal approach would be to only sample holotypes or paratypes, even if accessing these can be challenging as they are valuable to natural history museum collections and may not be readily available for DNA sampling if the integrity of morphological characters is jeopardized. In addition to taxonomic uncertainty, sampling locations and dates might be incomplete or unreliable. In consequence, label information on the specimen should always be considered with caution, as it can contains inaccurate or false information (Boessenkool et al. 2010, Rawlence et al. 2014, Palandačić et al. 2020)

An additional limitation is the scarcity of data on the curatorial handling of museum specimens, which is helpful for the application of museomics to wet collections. Information on critical parameters such as formalin concentration, duration of formalin exposure, pH levels and storage temperatures could help better understand and anticipate the potential degradation of DNA within these specimens, but this data is infrequently recorded. This also raises concerns about the potential contamination of museum specimens, which can occur due to various routines associated with accessing and manipulating museum collections, as well as the growth of bacteria or fungi over time.

Ethical challenges related to specimens' origins

It is important to acknowledge that many of the specimens in natural history museum collections were collected in locations outside the museum's country of origin. To address this, we should explore ways to meaningfully engage researchers from the countries where these specimens were sampled. This could involve collaborating with taxonomists and biodiversity experts from these countries to validate the identifications of taxa and facilitate their participation in the research process. Making the data and findings from museum-based studies openly accessible online would also enable equitable recognition and participation of researchers from the countries of specimen origin. Natural history museums have invested significant resources to digitise their collections and make them widely accessible through online platforms like GBIF. These collaborative data-sharing initiatives represent important steps towards a more inclusive approach to leveraging museum collections for research.

Conclusion

Natural history collections are inherently interdisciplinary, linking various fields such as taxonomy, genetics and conservation. Through both their collections and their role as scientific institutions, natural history museums have the potential to contribute to the advancement of eDNA research for biodiversity monitoring. They offer a vast resource of museum specimens that remain underutilised in genetic research and also have the potential to serve as repositories for freshly collected eDNA samples (de Santana et al. 2021). Many species without available DNA barcodes can be found in these collections, providing an opportunity to fill gaps in reference databases. Additionally, some parallels between working with fragmented DNA from museum specimens and eDNA can lead to transferable methodological advancements. Both sources frequently involve short and degraded DNA sequences. When appropriate genomic regions are targeted and sequenced, they can provide useful taxonomic information. Well-developed museomics methodologies, such as hybridisation-capture, could potentially be applied to eDNA to improve the recovery and sequencing of target taxa. Ultimately, the integration of museomics into eDNA research has the potential to enhance our understanding of global biodiversity, facilitate effective environmental monitoring, and support conservation effort.

Acknowledgements

We would like to thank Julia Bilat, Inés Carrasquer, Cody Cardenas, Nike Bahlmann, Nikola Djoric, Clare Holleley, Erin Hahn and Moritz Muschick for their precious advice on different museomics methods. We are grateful to Louise Lindblom for her valuable assistance in the lab, and to Ann-Helén Rønning and Min Chai for their help with the sampling. AI tools were used to improve the grammar and syntax of the manuscript. Part of this project was possible thanks to the award of the Scientific Exchange Grant (grant number 10549) by the European Molecular Biology Organization (EMBO). This work has been funded by the French National Research Agency, through Generic call (PRCI, French-Swiss call, ANR-21-CE02-0032, SHIFTeDNA) and by the Swiss National Science Foundation (grant number 205556).

Data availability

The raw sequencing reads were deposited in the following SRA repository http://www.ncbi.nlm.nih.gov/bioproject/1200624 (BioProject ID PRJNA1200624, release upon acceptance of the manuscript). Scripts to generate the datasets and plots are available on Dryad DOI: 10.5061/dryad.0zpc8677g (now private, but access can be provided to reviewers).

References

Agne S, Preick M, Straube N, Hofreiter M. 2022. Simultaneous barcode sequencing of diverse museum collection specimens using a mixed RNA bait set. *Frontiers in Ecology and Evolution* 10: 909846.

Allen MC, Kwait R, Vastano A, Kisurin A, Zoccolo I, Jaffe BD, Angle JC, Maslo B, Lockwood JL. 2023. Sampling environmental DNA from trees and soil to detect cryptic arboreal mammals. *Scientific Reports* 13: 180.

Andres KJ, Sethi SA, Lodge DM, Andrés J. 2021. Nuclear eDNA estimates population allele frequencies and abundance in experimental mesocosms and field samples. *Molecular Ecology* 30: 685–697.

Andres KJ, Lodge DM, Sethi SA, Andrés J. 2023. Detecting and analysing intraspecific genetic variation with eDNA: From population genetics to species abundance. *Molecular Ecology 32*: 4118–4132.

Aucone E, Kirchgeorg S, Valentini A, Pellissier L, Deiner K, Mintchev S. 2023. Drone-assisted collection of environmental DNA from tree branches for biodiversity monitoring. *Science Robotics* 8: eadd5762.

Baer J, Schliewen UK, Schedel FDB, Straube N, Roch S, Brinker A. 2023. Cryptic persistence and loss of local endemism in Lake Constance charr subject to anthropogenic disturbance. *Ecological Applications* 33: e2773.

Bakker J, Wangensteen OS, Chapman DD, Boussarie G, Buddo D, Guttridge TL, Hertler H, Mouillot D, Vigliola L, Mariani S. 2017. Environmental DNA reveals tropical shark diversity in contrasting levels of anthropogenic impact. *Scientific Reports* 7: 16886.

Bartonek JC, Nettleship DN, Natural Resources Council of America, U.S. Fish and Wildlife Service. 1979. *Conservation of Marine Birds of Northern North America : Papers from the International Symposium Held at the Seattle Hyatt House, Seattle, Washington, 13-15 May 1975*. Washington, D.C: U.S. Dept. of the Interior, Fish and Wildlife Service.

Bekaert B, Ellerington R, Van den Abbeele L, Decorte R. 2016. In-solution hybridization for the targeted enrichment of the whole mitochondrial genome. Pages 173–183 in Goodwin W, ed. *Forensic DNA typing protocols*. Springer.

Bemis KE, Girard MG, Santos MD, Carpenter KE, Deeds JR, Pitassy DE, Flores NAL, Hunter ES, Driskell AC, Macdonald KS, Weigt LA, Williams JT. 2023. Biodiversity of Philippine marine fishes: A DNA barcode reference library based on voucher specimens. *Scientific Data* 10: 411.

Beng KC, Corlett RT. 2020. Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. *Biodiversity and Conservation* 29: 2089–2121.

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. 2013. Unlocking the vault: next-generation museum population genomics. *Molecular Ecology* 22: 6018–6032.

Bi K, Linderoth T, Singhal S, Vanderpool D, Patton JL, Nielsen R, Moritz C, Good JM. 2019. Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLOS Genetics* 15: e1008119.

Blackman RC, Walser J-C, Rüber L, Brantschen J, Villalba S, Brodersen J, Seehausen O, Altermatt F. 2023. General principles for assignments of communities from eDNA: Open versus closed taxonomic databases. *Environmental DNA* 5: 326–342.

Boessenkool S, Star B, Scofield RP, Seddon PJ, Waters JM. 2010. Lost in translation or deliberate falsification? Genetic analyses reveal erroneous museum data for historic penguin specimens. *Proceedings. Biological Sciences* 277: 1057–1064.

Boettiger C, Lang DT, Wainwright PC. 2012. rfishbase: exploring, manipulating and visualizing FishBase data from R. *Journal of Fish Biology* 81: 2030–2039.

Boulanger E et al. 2021. Environmental DNA metabarcoding reveals and unpacks a biodiversity conservation paradox in Mediterranean marine reserves. *Proceedings of the Royal Society B: Biological Sciences* 288: rspb.2021.0112, 20210112.

Boussarie G et al. 2018. Environmental DNA illuminates the dark diversity of sharks. *Science Advances* 4: eaap9661.

Bradley RD, Bradley LC, Garner HJ, Baker RJ. 2014. Assessing the value of natural history collections and addressing issues regarding long-term growth and care. *BioScience* 64: 1150–1158.

Buckner JC, Sanders RC, Faircloth BC, Chakrabarty P. 2021. The critical importance of vouchers in genomics. *eLife* 10: e68264.

Burrell AS, Disotell TR, Bergey CM. 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* 79: 35–44.

Byerly PA, Chesser RT, Fleischer RC, McInerney N, Przelomska NAS, Leberg PL. 2022. Museum genomics provide evidence for persistent genetic differentiation in a threatened seabird species in the Western Atlantic. *Integrative and Comparative Biology* 62: 1838–1848.

Chang ES. 2020. Beyond specimens: Research, education, and policy. *BioScience* 70: 967–970.

Chevalier M, Broennimann O, Guisan A. 2024. Climate change may reveal currently unavailable parts of species' ecological niches. *Nature Ecology & Evolution* 8: 1298–1310.

Clare EL, Economou CK, Bennett FJ, Dyer CE, Adams K, McRobie B, Drinkwater R, Littlefair JE. 2022. Measuring biodiversity from DNA in the air. *Current Biology* 32: 693-700.e5.

Collins RA, Bakker J, Wangensteen OS, Soto AZ, Corrigan L, Sims DW, Genner MJ, Mariani S. 2019. Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution* 10: 1985–2001.

Cook JA et al. 2020. Integrating biodiversity infrastructure into pathogen discovery and mitigation of emerging infectious diseases. *BioScience* 70: 531–534.

Cook S, Dodge C, Morgan R, Sandusky GE. 2014. DNA/RNA degradation rate in long term fixed museum specimens. *Forensic Medicine and Anatomy Research* 03: 1.

Curd EE et al. 2019. Anacapa Toolkit: An environmental DNA toolkit for processing multilocus metabarcode datasets. *Methods in Ecology and Evolution* 10: 1469–1475.

Dabney J et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* 110: 15758–15763.

Davis CC, Knapp S. 2024. Exploring biodiversity through museomics. *Nature Reviews Genetics* 1–2.

De Bruyn M, Parenti LR, Carvalho GR. 2011. Successful extraction of DNA from archived alcohol-fixed white-eye fish specimens using an ancient DNA protocol. *Journal of Fish Biology* 78: 2074–2079.

Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters* 10: 20140562.

Dietz L et al. 2023. Standardized nuclear markers improve and homogenize species delimitation in Metazoa. *Methods in Ecology and Evolution* 14: 543–555.

Duval K, Aubin RA, Elliott J, Gorn-Hondermann I, Birnboim HC, Jonker D, Fourney RM, Frégeau CJ. 2010. Optimized manual and automated recovery of amplifiable DNA from tissues preserved in buffered formalin and alcohol-based fixative. *Forensic Science International: Genetics* 4: 80–88.

Dziedzic E, Sidlauskas B, Cronn R, Anthony J, Cornwell T, Friesen TA, Konstantinidis P, Penaluna BE, Stein S, Levi T. 2023. Creating, curating and evaluating a mitogenomic reference database to improve regional species identification using environmental DNA. *Molecular Ecology Resources* 23: 1880–1904.

Elith J et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.

Ellwood ER et al. 2020. Biodiversity science and the twenty-first century workforce. *BioScience* 70: 119–121.

Fong JJ, Blom MPK, Aowphol A, McGuire JA, Sutcharit C, Soltis PS. 2023. Editorial: Recent advances in museomics: revolutionizing biodiversity research. *Frontiers in Ecology and Evolution* 11.

Formenti G et al. 2022. The era of reference genomes in conservation genomics. *Trends in Ecology & Evolution* 37: 197–202.

Fraija-Fernández N, Bouquieaux M-C, Rey A, Mendibil I, Cotano U, Irigoien X, Santos M, Rodríguez-Ezpeleta N. 2020. Marine water environmental DNA metabarcoding provides a comprehensive fish diversity assessment and reveals spatial patterns in a large oceanic area. *Ecology and Evolution* 10: 7560–7584.

Funk DJ, Omland KE. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics* 34: 397–423.

Gansauge M-T, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* 8: 737–748.

Gansauge M-T, Aximu-Petri A, Nagel S, Meyer M. 2020. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nature Protocols* 15: 2279–2300.

García-Navas V, Sattler T, Schmid H, Ozgul A. 2020. Temporal homogenization of functional and beta diversity in bird communities of the Swiss Alps. *Diversity and Distributions* 26: 900–911.

Gasc C, Peyretaillade E, Peyret P. 2016. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic Acids Research* 44: 4504-4518.

Geburzi JC, Rodríguez-Flores PC, Derkarabetian S, Giribet G. 2024. From the shallows to the depths: a new probe set to target ultraconserved elements for Decapoda and other Malacostraca. *Frontiers in Marine Science* 11.

González Fortes G, Paijmans JLA. 2019. Whole-genome capture of ancient DNA using homemade baits. Pages 93–105 in Shapiro B, Barlow A, Heintzman PD, Hofreiter M, Paijmans JLA, and Soares AER, eds. *Ancient DNA: Methods and Protocols*. Springer.

Grinstein JD. 2023. When long synthetic DNA is needed in short order. *Genetic Engineering & Biotechnology* News 43: 20–23.

Günther B, Marre S, Defois C, Merzi T, Blanc P, Peyret P, Arnaud-Haond S. 2022. Capture by hybridization for full-length barcode-based eukaryotic and prokaryotic biodiversity inventories of deep sea ecosystems. *Molecular Ecology Resources* 22: 623–637.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucleic acids research* 41: e129–e129.

Hahn EE, Alexander MR, Grealy A, Stiller J, Gardiner DM, Holleley CE. 2022. Unlocking inaccessible historical genomes preserved in formalin. *Molecular Ecology Resources* 22: 2130–2147.

Haÿ V, Mennesson MI, Dettaï A, Bonillo C, Keith P, Lord C. 2020. Needlepoint non-destructive internal tissue sampling for precious fish specimens. *Philippe KeItH* 44: 73-79.

Heberling JM, Prather LA, Tonsor SJ. 2019. The changing uses of herbarium data in an era of global change: An overview using automated content analysis. *BioScience* 69: 812–822.

IUCN. 2024. The IUCN Red list of threatened species. Version 2024-2. (25 September 2024; *www.iucnredlist.org*).

Jones CB, Stock K, Perkins SE. 2024. AI-based discovery of habitats from museum collections. *Trends in Ecology & Evolution* 39: 323–327.

Kapp JD, Green RE, Shapiro B. 2021. A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *Journal of Heredity* 112: 241–249.

Keck F, Couton M, Altermatt F. 2022. Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Molecular Ecology Resources* 23: 742-755.

Krehenwinkel H et al. 2019. Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *GigaScience* 8: giz006.

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. TimeTree 5: An expanded resource for species divergence times. *Molecular Biology and Evolution* 39: msac174.

Lauridsen H, Pedersen JMH, Ringgaard S, Møller PR. 2022. Buoyancy and hydrostatic balance in a West Indian Ocean coelacanth *Latimeria chalumnae*. *BMC Biology* 20: 180.

Leerhøi F et al. 2024. Exploring the potential of extreme citizen science with Danish high school students using environmental DNA for marine monitoring. *Frontiers in Marine Science* 11.

Lemmon EM, Lemmon AR. 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44: 99–121.

Lenoir J, Bertrand R, Comte L, Bourgeaud L, Hattab T, Murienne J, Grenouillet G. 2020. Species better track climate warming in the oceans than on land. *Nature Ecology & Evolution* 4: 1044–1059.

Lentz DL, Hamilton TL, Dunning NP, Tepe EJ, Scarborough VL, Meyers SA, Grazioso L, Weiss AA. 2021. Environmental DNA reveals arboreal cityscapes at the Ancient Maya Center of Tikal. *Scientific Reports* 11: 12725.

Leray M, Knowlton N, Ho S-L, Nguyen BN, Machida RJ. 2019. GenBank is a reliable resource for 21st century biodiversity research. *Proceedings of the National Academy of Sciences* 116: 22651–22656.

Levesque-Beaudin V et al. 2023. A workflow for expanding DNA barcode reference libraries through 'museum harvesting' of natural history collections. *Biodiversity Data Journal* 11: e100677.

Li J, Seeber P, Axtner J, Crouthers R, Groenenberg M, Koehncke A, Courtiol A, Chanratana P, Greenwood AD. 2023. Monitoring terrestrial wildlife by combining hybridization capture and metabarcoding data from waterhole environmental DNA. *Biological Conservation* 284: 110168.

Locatelli NS, McIntyre PB, Therkildsen NO, Baetscher DS. 2020. GenBank's reliability is uncertain for biodiversity researchers seeking species-level assignment for eDNA. *Proceedings of the National Academy of Sciences* 117: 32211–32212.

Lyra ML, Lourenço ACC, Pinheiro PDP, Pezzuti TL, Baêta D, Barlow A, Hofreiter M, Pombal JP JR, Haddad CFB, Faivovich J. 2020. High-throughput DNA sequencing of museum specimens sheds light on the long-missing species of the *Bokermannohyla claresignata* group (Anura: Hylidae: Cophomantini). *Zoological Journal of the Linnean Society* 190: 1235–1255.

MacLean HJ, Nielsen ME, Kingsolver JG, Buckley LB. 2018. Using museum specimens to track morphological shifts through climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374: 20170404.

Mariac C, Vigouroux Y, Duponchelle F, García-Dávila C, Nunez J, Desmarais E, Renno JF. 2018. Metabarcoding by capture using a single COI probe (MCSP) to identify and quantify fish species in ichthyoplankton swarms. *PLOS ONE* 13: e0202976.

Marques V, Milhau T, Albouy C, Dejean T, Manel S, Mouillot D, Juhel J-B. 2021. GAPeDNA: Assessing and mapping global species gaps in genetic databases for eDNA metabarcoding. *Diversity and Distributions* 27: 1880–1892.

Mathon L et al. 2022. Cross-ocean patterns and processes in fish biodiversity on coral reefs through the lens of eDNA metabarcoding. *Proceedings. Biological Sciences* 289: 20220162.

McDonough MM, Parker LD, Rotzel McInerney N, Campana MG, Maldonado JE. 2018. Performance of commonly requested destructive museum samples for mammalian genomic studies. *Journal of Mammalogy* 99: 789–802.

McGuire JA et al. 2018. Squeezing water from a stone: high-throughput sequencing from a 145-year-old holotype resolves (barely) a cryptic species problem in flying lizards. *PeerJ* 6: e4470.

Meiklejohn KA, Damaso N, Robertson JM. 2019. Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. *PLOS ONE* 14: e0217084.

Meineke EK, Davies TJ, Daru BH, Davis CC. 2018. Biological collections for understanding biodiversity in the Anthropocene. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374: 20170386.

Miller SE et al. 2020. Building natural history collections for the twenty-first century and beyond. *BioScience* 70: 674–687.

Mollen FH, Iglésias SP. 2023. An inventory of Bramble sharks *Echinorhinus brucus* (Bonnaterre, 1788) (Elasmobranchii, Echinorhinidae) in natural history collections worldwide for conservation status assessment. *Zoosystema* 45.

Muschick M, Nikolaeva E, Rüber L, Matschiner M. 2022. The mitochondrial genome of the red icefish (*Channichthys rugosus*) casts doubt on its species status. *Polar Biology* 45: 1541-1552

Palandačić A, Kapun M, Greve C, Schell T, Kirchner S, Kruckenhauser L, Szucsich N, Bogutskaya N. 2024. From historical expedition diaries to whole genome sequencing: A case study of the likely extinct Red Sea torpedo ray. *Zoologica* Scripta 53: 32–51.

Palandačić A, Kruckenhauser L, Ahnelt H, Mikschi E. 2020. European minnows through time: museum collections aid genetic assessment of species introductions in freshwater fishes (Cyprinidae: *Phoxinus* species complex). *Heredity* 124: 410–422.

Pereira CL, Gilbert MTP, Araújo MB, Matias MG. 2021. Fine-tuning biodiversity assessments: A framework to pair eDNA metabarcoding and morphological approaches. *Methods in Ecology and Evolution* 12: 2397–2409.

Pereira HM et al. 2010. Scenarios for global biodiversity in the 21st century. *Science* 330: 1496–1501.

Piñol J, Mir G, Gomez-Polo P, Agustí N. 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources* 15: 819–830.

Ramírez-Amaro S, Bassitta M, Picornell A, Ramon C, Terrasa B. 2022. Environmental DNA: State-of-the-art of its application for fisheries assessment in marine environments. *Frontiers in Marine Science* 9: 1004674.

Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes* 7: 355–364.

Rawlence N, Kennedy M, Waters J, Scofield R. 2014. Morphological and ancient DNA analyses reveal inaccurate labels on two of Buller's bird specimens. *Journal of the Royal Society of New Zealand* 44: 163–169.

Raxworthy CJ, Smith BT. 2021. Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution* 36: 1049–1060.

Renner SS, Scherz MD, Schoch CL, Gottschling M, Vences M. 2024. Improving the gold standard in NCBI GenBank and related databases: DNA sequences from type specimens and type strains. *Systematic Biology* 73: 486–494.

Rozanski R, Trenkel VM, Lorance P, Valentini A, Dejean T, Pellissier L, Eme D, Albouy C. 2022. Disentangling the components of coastal fish biodiversity in southern Brittany by applying an environmental DNA approach. *Environmental DNA* 4: 920–939.

Ruane S, Austin CC. 2017. Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. *Molecular Ecology Resources* 17: 1003–1008.

de Santana CD et al. 2021. The critical role of natural history museums in advancing eDNA for biodiversity studies: a case study with Amazonian fishes. *Scientific Reports* 11: 18159.

Schmid S, Genevest R, Gobet E, Suchan T, Sperisen C, Tinner W, Alvarez N. 2017. HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. *Methods in Ecology and Evolution* 8: 1374–1388.

Seeber PA, McEwen GK, Löber U, Förster DW, East ML, Melzheimer J, Greenwood AD. 2019. Terrestrial mammal surveillance using hybridization capture of environmental DNA from African waterholes. *Molecular Ecology Resources* 19: 1486–1496.

Seymour M, Edwards FK, Cosby BJ, Kelly MG, de Bruyn M, Carvalho GR, Creer S. 2020. Executing multi-taxa eDNA ecological assessment via traditional metrics and interactive networks. *Science of The Total Environment* 729: 138801.

Simmons JE. 2014. *Fluid Preservation*. Rowman & Littlefield Publishers.

Staudinger MD et al. 2013. Biodiversity in a changing climate: a synthesis of current and projected trends in the US. *Frontiers in Ecology and the Environment* 11: 465–473.

Straube N et al. 2021. Successful application of ancient DNA extraction and library construction protocols to museum wet collection specimens. *Molecular Ecology Resources* 21: 2299–2315.

Suchan T, Pitteloud C, Gerasimova NS, Kostikova A, Schmid S, Arrigo N, Pajkovic M, Ronikier M, Alvarez N. 2016. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLOS ONE* 11: e0151651.

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. 2012. Environmental DNA. *Molecular Ecology* 21: 1789–1793.

Valentini A et al. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* 25: 929–942.

Veron P et al. 2023. Environmental DNA complements scientific trawling in surveys of marine fish biodiversity. *ICES Journal of Marine Science* 80: 2150–2165.

Wales N, Carøe C, Sandoval-Velasco M, Gamba C, Barnett R, Samaniego JA, Ramos Madrigal J, Orlando L, Gilbert MTP. 2015. New insights on single-stranded versus double-stranded dna library preparation for ancient DNA. *BioTechniques* 59: 368–371.

Walsh HJ, Richardson DE, Marancik KE, Hare JA. 2015. Long-term changes in the distributions of larval and adult fish in the northeast U.S. shelf ecosystem. *PLOS ONE* 10: e0137382.

Wandeler P, Hoeck PEA, Keller LF. 2007. Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution* 22: 634–642.

Wang Z, Liu X, Liang D, Wang Q, Zhang L, Zhang P. 2023. VertU: universal multilocus primer sets for eDNA metabarcoding of vertebrate diversity, evaluated by both artificial and natural cases. *Frontiers in Ecology and Evolution* 11: 1164206.

Wilcox TM, Zarn KE, Piggott MP, Young MK, McKelvey KS, Schwartz MK. 2018. Capture enrichment of aquatic environmental DNA: A first proof of concept. *Molecular Ecology Resources*. 18: 1392-1401.

Winker K. 2004. Natural history museums in a postbiodiversity era. *BioScience* 54: 455–459.

Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15: R46.

Wood HM, González VL, Lloyd M, Coddington J, Scharff N. 2018. Next-generation museum genomics: Phylogenetic relationships among palpimanoid spiders using sequence capture techniques (Araneae: Palpimanoidea). *Molecular Phylogenetics and Evolution* 127: 907–918.

Yang C, Bohmann K, Wang X, Cai W, Wales N, Ding Z, Gopalakrishnan S, Yu DW. 2021. Biodiversity Soup II: A bulk-sample metabarcoding pipeline emphasizing error reduction. *Methods in Ecology and Evolution* 12: 1252–1264.

Zhu T, Sato Y, Sado T, Miya M, Iwasaki W. 2023. MitoFish, MitoAnnotator, and MiFish Pipeline: Updates in 10 Years. *Molecular Biology and Evolution* 40: msad035.

Webster MS. 2017. *The Extended Specimen: Emerging Frontiers in Collections-Based Ornithological Research*. CRC Press.

Supplemental figures and material and methods for "Combining eDNA and museomics to enhance biodiversity monitoring"

Sarah Schmid, Nicolas Straube, Camille Albouy, Bo Delling, James Maclaine, Michael Matschiner, Peter Rask Møller, Annamaria Nocita, Anja Palandačić, Lukas Rüber, Moritz Sonnewald, Nadir Alvarez, Stéphanie Manel, Loïc Pellissier

Content

Supplemental tables

Table S1. List of downloaded datasets from European natural history museum collection available on GBIF on the 4th of June 2024.

DOI	Collection	City	Taxons	Citation
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.kq3335	Museu de Ciències Naturals de Barcelona: MCNB-Cord	Barcelona	chordata	Quesada Lara J, Agulló Villaronga J (2024). Museu de Ciències Naturals de Barcelona: MCNB-Cord. Museu de Ciències Naturals de Barcelona. Occurrence dataset https://doi.org/10.15468/yta7zj accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.mye954	Fish Collection NRM, Swedish Museum of Natural History	Stockholm	fish	Delling B, Holston K (2024). Fish Collection NRM. Version 43.430. Swedish Museum of Natural History. Occurrence dataset https://doi.org/10.15468/d7eitf accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.yys7ar	RBINS Fish collection	Bruxelles	fish	Pauwels O, Vandenberghe T, Cooleman S, Theeten F (2023). RBINS Fish collection. Version 1.12. Royal Belgian Institute of Natural Sciences. Occurrence dataset https://doi.org/10.15468/ga5ady accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.yxjtbj	Zoological collection of the National Museum	Prague	chordata	National Museum in Prague. Zoological collection of the National Museum. Occurrence dataset https://doi.org/10.15468/4yiuil accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.2wpvmp	NHMD Ichthyology Collection	Copenhagen	fish	Møller P R, Carl H (2024). NHMD Ichthyology Collection. Natural History Museum of Denmark. Occurrence dataset https://doi.org/10.15468/cs1ywg accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.j3faqs	Estonian Museum of Natural History Department of Zoology	Talinn	chordata	Lennuk L. Estonian Museum of Natural History Department of Zoology. Estonian Museum of Natural History. Occurrence dataset https://doi.org/10.15468/98cxtc accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.g2crgg	The fishes collection (IC) of the Muséum national d'Histoire naturelle (MNHN - Paris)	Paris	fish	MNHN, Chagnoux S (2024). The fishes collection (IC) of the Muséum national d'Histoire naturelle (MNHN - Paris). Version 57.364. MNHN - Museum national d'Histoire naturelle. Occurrence dataset https://doi.org/10.15468/tm7whu accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.kce3ag	ZFMK Ichthyology collection	Bonn	fish	Leibniz Institute for the Analysis of Biodiversity Change (LIB). ZFMK Ichthyology collection. Occurrence dataset https://doi.org/10.15468/nvuhqv accessed via GBIF.org on 2024-06-04.

GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.2nyewr	MfN Fish Collection, Zoology	Berlin	fish	Museum für Naturkunde Berlin. MfN Fish Collection, Zoology. Occurrence dataset https://doi.org/10.15468/bq2dwf accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.gpfnwy	Collection Pisces SMF	Frankfurt (Senckenberg)	fish	Senckenberg. Collection Pisces SMF. Occurrence dataset https://doi.org/10.15468/xaofbe accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.za8z7p	Collection Ichthyologie - SNSD	Frankfurt (Senckenberg)	fish	Senckenberg. Collection Ichthyologie - SNSD. Occurrence dataset https://doi.org/10.15468/wnwumq accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.fy47fr	Naturalis Biodiversity Center (NL) - Pisces	Leiden	fish	Dondorp E, Creuwels J (2024). Naturalis Biodiversity Center (NL) - Pisces. Naturalis Biodiversity Center. Occurrence dataset https://doi.org/10.15468/evijly accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.ebjjwy	Ichthyology Collection, University Museum Bergen, University of Bergen	Bergen	fish	University of Bergen (2024). Ichthyology Collection, University Museum Bergen, University of Bergen. Version 1.4. Occurrence dataset https://doi.org/10.15468/kvr5bz accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.nc7jvf	Fish collection (TSZP) The Arctic University Museum of Norway	Tromsø	fish	Frafjord K, Bergersen R, Altenburger A (2023). Fish collection (TSZP) The Arctic University Museum of Norway. Version 1.7. UiT The Arctic University of Norway. Occurrence dataset https://doi.org/10.15468/hdpah5 accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.48gqxf	Fish collection, Natural History Museum, University of Oslo	Oslo	fish	University of Oslo (2021). Fish collection, Natural History Museum, University of Oslo. Version 1.181. Occurrence dataset https://doi.org/10.15468/4vqytb accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.dfpbag	Museo Nacional de Ciencias Naturales, Madrid: MNCN_ICTIO	Madrid	fish	Solís Fraile G (2020). Museo Nacional de Ciencias Naturales, Madrid: MNCN_ICTIO. Museo Nacional de Ciencias Naturales (CSIC). Occurrence dataset https://doi.org/10.15468/emdpoi accessed via GBIF.org on 2024-06-04.
GBIF.org (04 June 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.cpq6d3	Natural History Museum (London) Collection Specimens	London	chordata	Natural History Museum (2024). Natural History Museum (London) Collection Specimens. Occurrence dataset https://doi.org/10.5519/qd.9vixygzc accessed via GBIF.org on 2024-06-04.

Table S2. List of museum specimens sequenced. Sequencing ID, species name, specimen catalog number, collection date, location date, host museum collection and tissue sampled for molecular analysis are indicated. Sequencing information is also included: the median read length [bp] custom/kit column indicates the median length of reads in base pair after filtering and trimming and the percentage of duplicated reads indicates the number of PCR duplicates for each sample. The first value is using the custom protocol from Straube et al. 2021, and the second value is using the "kit" protocol. In addition, information about the closest and "equally distant" reference genomes used for the Fastqscreen analysis is displayed as well as the divergence in million years (MYA) from the given reference genomes.

Sequencing ID	Target species	Specimen catalog number	Samping date	Sampling location	Museum collection	Tissue for DNA extraction	Median read length [bp] custom/kit	% of duplicated reads custom/kit	Closest reference whole-genome	Accession whole-genome	Distance to closer whole-genome	Equally distant genome	Accession equally distance genome	Distance to equally distant genome
117_STK	*Pseudobatos leucorhynchus*	NRM 9050	1852	Panama	Naturhistoriska riksmuseet, Stockholm, Sweden (NRM)	gut lining	20/24	74.75/26.6	*Rhynchobatus australiae*	GCA_034780895.1	86 MYA	*Mobula birostris*	GCA_030035685.1	165 MYA
185_OSL	*Lycodes palearis*	NHMO J1678	1891	Aleutian islands (USA)	Natural History Museum, Oslo, Norway (NHMO)	muscle	34/34	12.38/6.04	*Lycodopsis pacificus*	GCA_028022725.1	6.0 MYA	*Thaleichthys pacificus*	GCA_023658055.1	169 MYA
308_FIR	*Negaprion acutidens*	MZUF 2276	1973	Habo lagune (Somalia)	Museo Zoologico La Specola, Università di Firenze, Italy (MZUF)	muscle	38/74	10.43/1.8	*Negaprion brevirostris*	GCA_030324005.1	14 MYA	*Carcharodon carcharias*	GCA_017639515.1	173 MYA
360_WN	*Carcharhinus cerdale*	NMW 84790	1902	Panama	Naturhistoriches Museum Wien, Austria (NMW)	gill filaments	26/34	2.51/4.06	*Carcharhinus longimanus*	GCA_030264375.1	45 MYA	*Carcharodon carcharias*	GCA_017639515.1	173 MYA

472_WN	*Hydrolagus purpurescens*	NMW 50253	1905	Oshima (Japan)	Naturhistoriches Museum Wien, Austria (NMW)	muscle	16/74	51.96/32.94	*Hydrolagus affinis*	GCA_012026655.1	93 MYA	*Callorhinchus milii*	GCA_018977255.1	187 MYA
600_CP	*Ogilbia mccoskeri*	ZMUC P771387	1976	Toro point (Panama)	Zoological Museum, University of Copenhagen, Denmark (ZMUC)	gill filaments	40/70	1.82/2.22	*Lucifuga dentata*	GCA_014773175.1	44 MYA	*Thaleichthys pacificus*	GCA_023658055.1	169 MYA
612_CP	*Diancistrus katrineae*	ZMUC P771480	1971	Rarotonga (Cook Islands)	Zoological Museum, University of Copenhagen, Denmark (ZMUC)	gill filaments	28/24	4.24/11.53	*Lucifuga dentata*	GCA_014773175.1	44 MYA	*Thaleichthys pacificus*	GCA_023658055.1	169 MYA
648_CP	*Neoharriotta pumila*	ZMUC P0950	1988	Socotra island (Yemen)	Zoological Museum, University of Copenhagen, Denmark (ZMUC)	muscle	22/24	5.41/31.15	*Hydrolagus affinis*	GCA_012026655.1	122 MYA	*Callorhinchus milii*	GCA_018977255.1	187 MYA
660_CP	*Diancistrus tongaensis*	ZMUC P771489	1993	Vava'u islands (Tonga)	Zoological Museum, University of Copenhagen, Denmark (ZMUC)	muscle	16/74	54.25/26.81	*Lucifuga dentata*	GCA_014773175.1	44 MYA	*Thaleichthys pacificus*	GCA_023658055.1	169 MYA
418_WN	*Echinorhinus brucus*	NMW 85158	1910	Nice (France)	Naturhistoriches Museum Wien, Austria (NMW)	gills	32/40	6.77/6.36	*Squatina squatina*	GCA_031763465.1	181 MYA	*Squatina squatina*	GCA_031763465.1	181 MYA

Supplemental figures

Figure S1

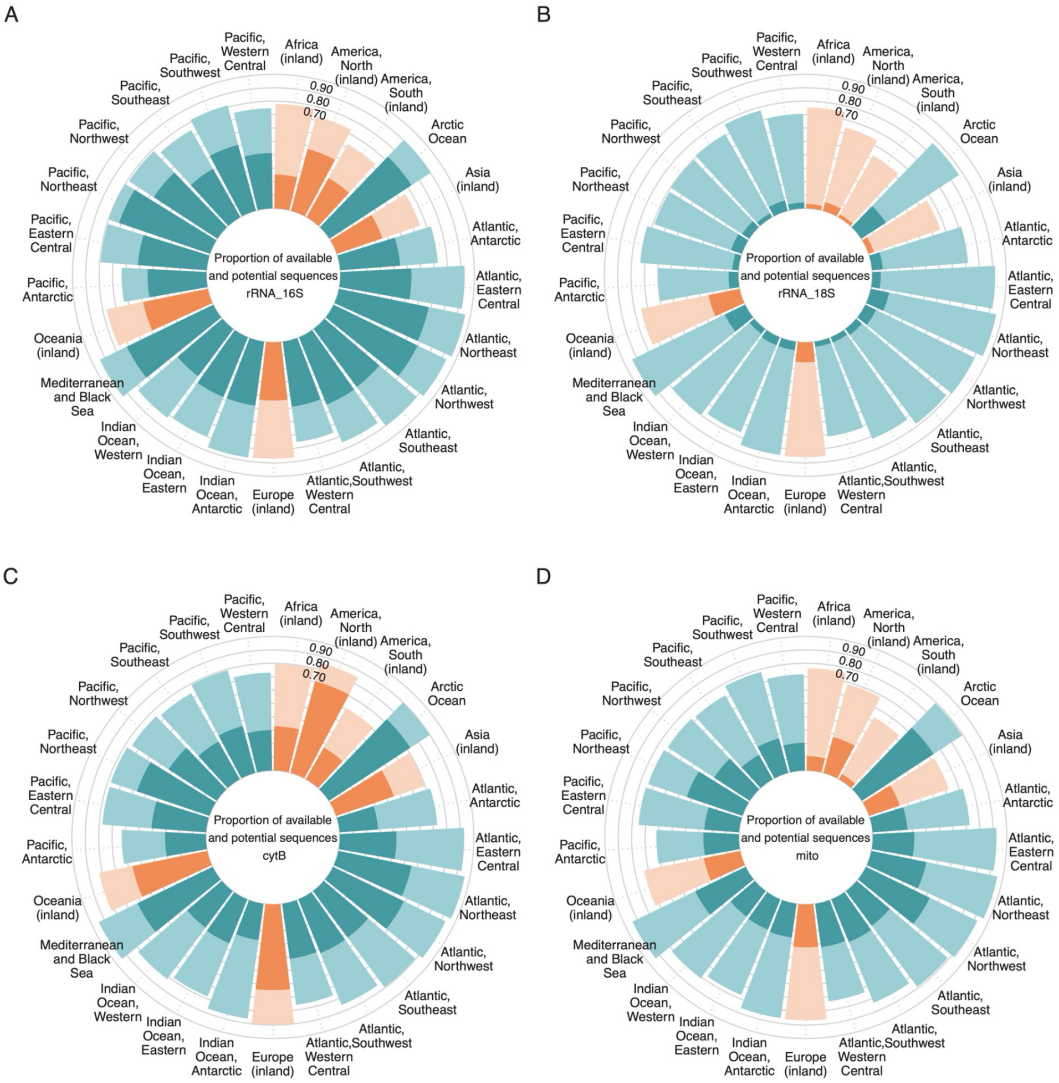

Figure S1. Potential of museum specimens to improve reference databases for 16S ribosomal RNA (16S) (A), 18S ribosomal RNA (18S) (B) and cytochrome B (cytB) (C) mitochondrial genes used in eDNA studies and for the full mitochondrial genome (D) according to geographic regions. Regions in blue are hosting marine species while regions in orange are for freshwater species (inland regions). Proportion of sequences already available is highlighted in dark colour. Potential new sequences based on available museum specimens in European collections are displayed in lighter colour. Potential of improvement according to geographic regions. The complete list of fish species was retrieved from FishBase using the R package rfishbase (Boettiger et al. 2012) as well as their corresponding geographic distribution. The museum data were retrieved from GBIF.

Figure S2

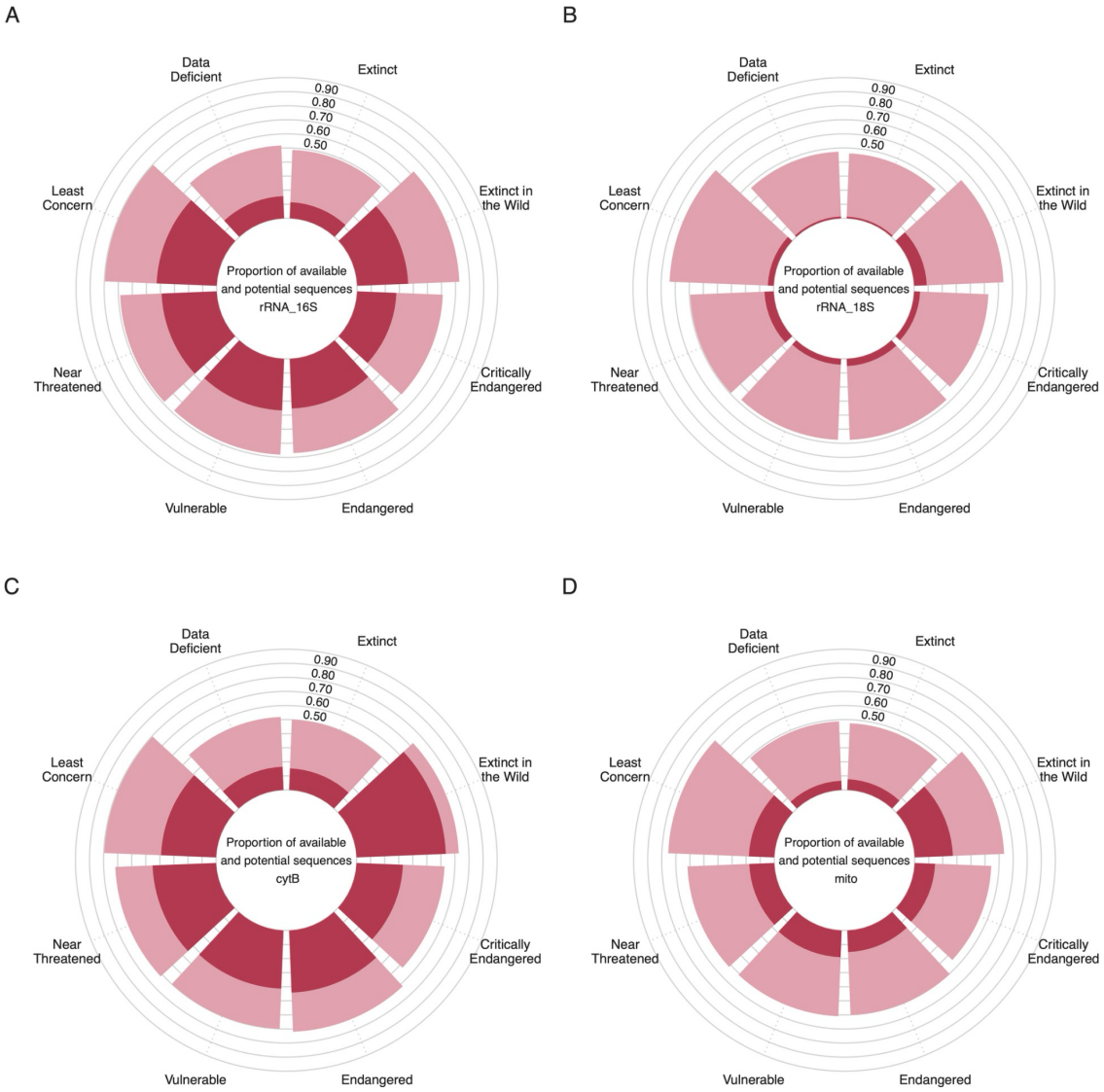

Figure S2. Potential of museum specimens to improve reference databases for for 16S ribosomal RNA (16S) (A), 18S ribosomal RNA (18S) (B) and cytochrome B (cytB) (C)mitochondrial genes used in eDNA studies and for the full mitochondrial genome (D) according to global IUCN Red List of Threatened Species assessments. Proportion of sequences already available is highlighted in dark colour. Potential new sequences based on available museum specimens in European collections are displayed in lighter colour. The complete list of fish species was retrieved from FishBase using the R package rfishbase (Boettiger et al. 2012). The museum data were retrieved from GBIF. The IUCN Red List assessments were retrieved from the IUCN website.

Figure S3

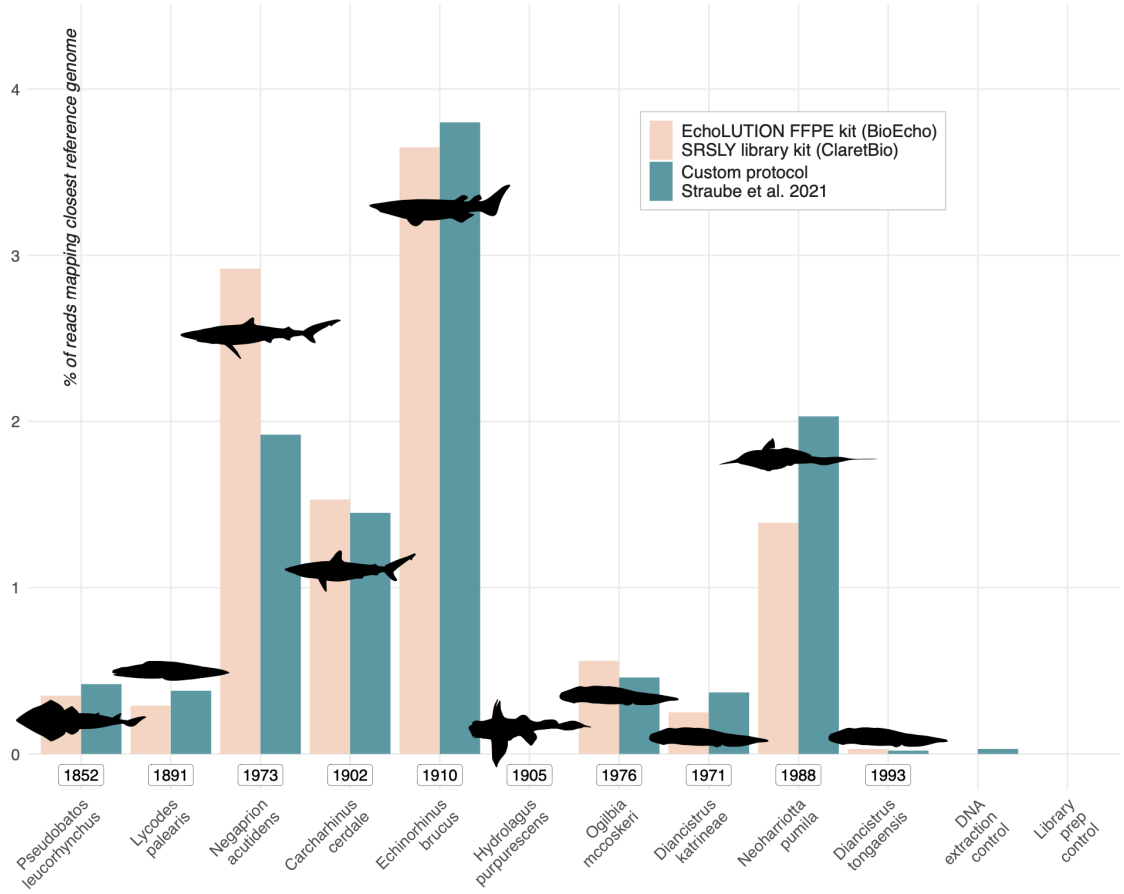

Figure S3. Percentage of reads mapping to the reference genome. Here, divergence between target species and reference genome for the ten species is approximately 180 MYA (same target-reference divergence for all species, as opposed to figure 3). This divergence corresponds to the maximal value between a target species (*Echinorhinus brucus*) and its closest available reference genome (*Squatina squatina*). Light pink bars highlight results based on the "commercial" protocol and blue bars are for the "custom" protocol. The squared value below each pair of bar plot correspond to the collection date of the museum specimens.

.

Supplemental material and methods

Data retrieval and comparison of eDNA barcodes sequences with museum specimen records

To examine the cumulative number of species sequenced for a given DNA barcode/mitochondrial genome (also referred as mitogenome) over the years, we retrieved all data available from NCBI using the R package rentrez v1.2.3 (Winter 2017). We searched the nucleotide database for the rRNA 12S, rRNA 16S, rRNA 18S, cytochrome B (cytB), cytochrome oxydase I (COI) barcodes as well as for the complete mitogenomes for all fish orders. In addition, we also retrieved all the fish species with available data on the sequence read archive (SRA) using the Entrez Direct (Kans 2024), which provides access to the NCBI databases from a Unix terminal window.

To highlight the potential of museum specimens for increasing the number of species with an available barcode/mitogenome sequence, we first downloaded all available datasets on the Global Biodiversity Information Facility (GBIF) listing fish specimens stored in European natural history museum collection (see table S1). Subsequently, we downloaded a list of all existing fish species using the R package rfishbase v5.0.0 (Boettiger et al. 2012) and extracted their geographic range (field AreaCode). In addition, we retrieved information about the Red List status of all fish species from the International Union for Conservation of Nature and Natural Resources (IUCN) website. All the datasets (barcodes, museum specimens, IUCN status and fish species list, and geographic range) were combined in R v.4.3.0 and subsequently plotted. All the scripts to retrieve data and to generate the figures are available on Dryad (DOI: 10.5061/dryad.0zpc8677g).

Specimens sampling, DNA extraction, library preparation and sequencing

To compare the effectiveness of commercial kits for DNA extraction and library preparation with a customised museomics protocol to retrieve genomic information from fish specimens from wet museum collections, we sampled ten distinct fish species in different European museum collections (table S2). The specimens were originally collected between 1852 and 1993 in various locations, and their preservation history is mostly unknown (table S2). However, specimens collected after 1910 were likely fixed with formaldehyde before long-term storage in ethanol.

For the “commercial” protocol, we extracted DNA with the EchoLUTION FFPE DNA Kit (BioEcho Life Sciences, Cologne, Germany). Since the protocol is originally made for paraffin-embedded blocks of tissue, we adapted the manufacturer protocol to fit it to potentially formalin-fixed wet-collection specimens. We washed the tissue samples with 2 mL PBS, cut it in small pieces with a scalpel blade and used between 5-15 ng tissue as input. We dropped the paraffin removal step, prolonged the decrosslinking steps to 50 minutes, increased lysis time to 1 hour and retrieved 80 µL of eluted DNA. After DNA extraction, we followed the SRSLY® PicoPlus Uracil+ kit protocol (Claret Bioscience, Scotts Valley, USA) to construct single-stranded DNA libraries. We proceeded to ten indexing PCR cycles and eluted the final DNA libraries in 20µL low TE (0.1M). We pooled all libraries together in equimolar ratio to reach a 4 nM molarity.

For the “custom” protocol, we followed the ‘Guanidine treatment’ DNA extraction from Straube et al. (2021), which is based on ancient DNA protocols and adapted to museum specimens from wet collections. In brief, after washing the tissues with 2 mL PBS and cutting them in small pieces, we incubated them in a guanidinium thiocyanate buffer for ~18 hours at 37°C. For meaningful comparisons, the same amount of tissue was used as for the “commercial” protocol. Subsequent DNA extraction steps follow Dabney et al. (2013) and (Rohland et al. 2004) procedures for ancient DNA extraction. We eluted the final purified DNA in twice 12.5 µL TET buffer (10 mM Tris-HCL, 1 mM EDTA, 0.05% Tween 20). Then, we prepared single-stranded libraries following (Gansauge et al. 2017) and determined the adequate number of indexing PCR cycles based on the mean cycle threshold (Ct) estimated by qPCR. We eluted the final DNA libraries in twice 10 µL of EB buffer.

We prepared two final library pools at 4 nM, pooling separately the libraries prepared with the commercial protocol and the libraries generated with the custom protocol. We sequenced each pool on an Illumina MiniSeq lane and generated 75 base pair-long single-end reads. For the custom protocol, we used custom read 1 and index 2 primers as described in Paijmans et al. 2017 (arXiv:1711.11004, preprint, not peer-reviewed).

Sequencing data preprocessing and mapping to reference genome

We trimmed the generated raw reads for adapter sequences as well as for low quality (below 10) nucleotides at both 5’ and 3’ ends using Cutadapt v.4.7 (Martin 2011). We assessed read quality before and after processing with FastQC v.0.11.7 (Andrews 2010). Then, to estimate the success in sequencing endogenous DNA for the ten fish species with the two different protocols, we used Fastq Screen (Wingett and Andrews 2018), which enables querying and subsampling large read datasets against a panel of reference genomes to identify the source of

the DNA sequences in our data. First, we mapped the reads of each species to the closest reference whole-genome available (table S2), estimated as the reference species with the lowest divergence time to each sample according to TimeTree (Kumar et al. 2022). Then, to evaluate the effect of genetic distance from the reference genome, we selected reference genomes that diverged approximately 180 million years ago from each target species, as this distance corresponds to the highest divergence observed between one of our focal species (*Echinorhinus brucus*) and the closest available reference genome (*Squatina squatina*).

References

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

Boettiger C, Lang DT, Wainwright PC. 2012. rfishbase: exploring, manipulating and visualizing FishBase data from R. *Journal of Fish Biology* 81: 2030–2039.

Gansauge M-T, Gerber T, Glocke I, Korlević P, Lippik L, Nagel S, Riehl LM, Schmidt A, Meyer M. 2017. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Research* 45: e79.

Kans J. 2024. *Entrez direct: E-utilities on the UNIX command line*. In Entrez programming utilities help [Internet]. National Center for Biotechnology Information (US).

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. TimeTree 5: An expanded resource for species divergence times. *Molecular Biology and Evolution* 39: msac174.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17: 10–12.

Rohland N, Siedel H, Hofreiter M. 2004. Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. *BioTechniques* 36: 814–821.

Straube N et al. 2021. Successful application of ancient DNA extraction and library construction protocols to museum wet collection specimens. *Molecular Ecology Resources* 21: 2299–2315.

Wingett SW, Andrews S. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* 7.

Winter DJ. 2017. rentrez: An R package for the NCBI eUtils API. *The R Journal* 9: 520–526.