Received: 18 March 2020 | Revised: 29 August 2020 | Accepted: 1 September 2020

DOI: 10.1002/edn3.138

ORIGINAL ARTICLE

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

WILEY

Mitochondrial genomes of Danish vertebrate species generated for the national DNA reference database, DNAmark

Ashot Margaryan[1,2,3] | Christina Lehmkuhl Noer[1,2] | Stine Raith Richter[1] | Marlene Elise Restrup[1] | Julie Lee Bülow-Hansen[4] | Frederik Leerhøi[1] | Emilia Marie Rolander Langkjær[1] | Shyam Gopalakrishnan[1,2] | Christian Carøe[1] | M. Thomas P. Gilbert[1,2,5] | Kristine Bohmann[1]

[1]Section for Evolutionary Genomics, Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[2]Center for Evolutionary Hologenomics, University of Copenhagen, Copenhagen, Denmark

[3]Institute of Molecular Biology, National Academy of Sciences, Yerevan, Armenia

[4]Department of Biology, Faculty of Science, University of Copenhagen, Copenhagen, Denmark

[5]Department of Natural History, NTNU, Trondheim, Norway

Correspondence
Ashot Margaryan, Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, 1353 Copenhagen K, Denmark.
Email: ashot.margaryan@sund.ku.dk

Funding information
Aage V. Jensen Naturfond

Abstract

Biodiversity monitoring projects using environmental DNA techniques are becoming increasingly widespread. However, these techniques depend heavily on the quality and richness of the available DNA reference database against which the DNA sequences are queried. To create a comprehensive DNA sequence database for future DNA-based biodiversity assessments in Denmark, a national DNA reference database, DNAmark, was established, which contains organellar and/or nuclear reference data from vouchered museum species of plants, animals, and fungi from Denmark. Here, we present full or partial mitochondrial genomes of 182 Danish vertebrate species representing ca. 22% of vertebrate species observed in Denmark. Further, we demonstrate that storage conditions of the specimens accounted for ca. 50% of the total variation in mitochondrial DNA (mtDNA) preservation while the age of museum specimens had little effect: ca. 4%. In addition, we roughly estimate the cost of sequencing to be 25 EUR per specimen for obtaining sufficient amounts of DNA reads (ca. 200-fold coverage) for reliable mitogenome assemblies while also obtaining low coverage genomic data. The large number of mitogenomes of Danish vertebrate species represents the initial groundwork for DNA-based biodiversity assessments of vertebrates in Denmark and paves the way for practitioners to freely choose mitochondrial DNA markers.

KEYWORDS

biodiversity assessment, COI, DNA barcoding, DNA metabarcoding, DNA reference databases, environmental DNA, genome skimming, mitochondrial DNA, molecular biodiversity assessment, next-generation sequencing, taxonomic identification

1 | INTRODUCTION

In recent years, molecular analyses of DNA extracted from environmental and bulk specimen samples have become a valuable tool in studies on biodiversity, diet, and ecological interactions (reviewed in Alberdi et al., 2018; Bohmann et al., 2014; Taberlet et al., 2012). The currently most popular approach to achieve DNA sequence-based taxonomic identifications of taxa within such samples is DNA metabarcoding. Metabarcoding relies on PCR amplification with metabarcoding primers targeting a taxonomically informative (for animals, mitochondrial) DNA marker sequence within a selected taxonomic group (Taberlet, Coissac, Pompanon, Coissac, Pompanon, Brochmann, & Willerslev, 2012). Following sequencing, the DNA marker sequences are compared with DNA reference databases to achieve taxonomic identification of sample constituents. For identification of animal sequences, there are two main DNA reference database options; NCBI GenBank and the Barcode of Life Data Systems (BOLD). The NCBI Genbank database is a large international DNA reference database that for animals contains annotated nuclear and mitochondrial sequences (ncbi.nlm.nih.gov/genbank, Benson et al., 2018). NCBI Genbank is considered a reliable resource for biodiversity research (Leray, Knowlton, Ho, Nguyen, & Machida, 2019; Meiklejohn, Damaso, & Robertson, 2019), but have also been reported to contain sequencing errors (Fietz, Graves, & Olsen, 2013). The BOLD database (boldsystems.org, Ratnasingham & Hebert, 2007) is another large, international DNA reference database, but in contrast to GenBank, it is based on taxonomically verified voucher specimens. Further, specimens are generally represented by the genetic region traditionally assigned to be the DNA barcode. For animals, this region is a 658 base pair (bp) cytochrome c oxidase subunit c (COI) marker (Hebert, Ratnasingham, & deWaard, 2003).

The barcode COI region has unfortunately been shown not to be an ideal universal marker region for metabarcoding studies of animals as primer binding sites within the protein-coding region are not highly conserved (Clarke, Soubrier, Weyrich, & Cooper, 2014; Deagle, Jarman, Coissac, Pompanon, & Tab erlet, 2014; Miya et al., 2015). This lack of conserved primer regions is due to synonymous mutations that do not change the coded COI protein (Deagle et al., 2014). This makes it hard to design primers that are truly conserved for a taxonomic group, which can cause failure to PCR amplify some taxa and lead to false negatives. However, since COI is the most represented genetic region found in taxonomically verified databases, metabarcoding studies of animals are caught in between the devil and the deep blue sea; if they change marker region to for example 12S or 16S to allow better amplification of taxa, then they will not be able to utilize the massive scale vouchered BOLD reference database and their taxonomic identifications will suffer. If they, on the other hand, stay with the COI as a marker, they lose reliability of PCR amplifying taxa, but the taxa they do amplify and sequence will stand a better chance at getting taxonomically identified.

To allow practitioners to change metabarcoding marker regions and pave the way for future environmental DNA approaches, several reference database projects have emerged that generate comprehensive reference data per specimen through genome skimming, that is, low coverage shotgun sequencing of total DNA extracted from each specimen (Coissac, Hollingsworth, Lavergne, & Taberlet, 2016). This approach allows simultaneous sequencing of different barcode markers and even assembly of mitochondrial genomes, and for plants chloroplast genomes and nuclear ribosomal DNAs, as well as recovery of low coverage nuclear data. For plants, the PhyloAlps and NorBOL (Norwegian initiative for the Barcoding of Life) projects generate plastid genomes and assemblies of nuclear ribosomal DNA for ca. 4,600 specimens representing the entire Alpine flora and for ca. 2000 specimens of vascular plants covering the arctic-boreal flora, respectively (Alsos et al., 2020). For vertebrate species, to our knowledge, there are no similar reference database projects based on genome skimming data. However, the ambitious Vertebrates Genome Project was initiated in 2017 as part of the Genome 10K project (https://genome10k.soe.ucsc.edu/) with the aim to generate reference genome assemblies of all ca. 66,000 vertebrate species; though, the project is far from being complete (https://vertebrategenomesproject.org/phase-one).

In Denmark, DNA-based biodiversity assessments are gaining increased foothold (e.g., Agersnap et al., 2017; Foote et al., 2012; Sigsgaard et al., 2017; Thomsen et al., 2012). Yet, as in other parts of the world efforts are restricted by the lack of reference data (e.g., Thomsen & Sigsgaard, 2019). A total of 834 vertebrate species have been observed in Denmark; these fall in the taxonomic orders Aves ($n = 477$), Pisces ($n = 241$), Mammalia ($n = 91$), Amphibia ($n = 15$), and Reptilia ($n = 10$) (allearter.dk). Of the 834 vertebrate species observed in Denmark, 373 species have complete mitogenomes available in the NCBI GenBank reference database. To improve DNA-based biodiversity assessments, there is therefore a need to develop Danish vertebrate DNA reference data, both with regard to the number of species and to the amount of reference data per species. To meet these needs, the national DNA reference database for Danish species, DNAmark, was established in 2017. In the DNAmark database, reference data are created through genome skimming of vouchered specimens of Danish species of plants, animals, and fungi. For animals, we use the sequence data to generate partial or full mitochondrial genomes, thereby covering mitochondrial markers typically used in environmental DNA studies, for example 12S, 16S, and COI. Such large mitochondrial assemblies provide additional scaffolds (in addition to the well-known marker regions) for species identification in environmental samples and better reference data for designing species-specific qPCR probes and primers. Moreover, mitochondrial genomes have more informative sites than, for example, the shorter COI region, and can therefore provide higher resolution in phylogenetic analyses.

Here, we generate and present the full or partial mitochondrial reference genomes generated in the DNAmark project for 192 vouchered specimens of 182 Danish vertebrate species spanning birds, fish, mammals, amphibians, and reptiles. Further, to guide future efforts to generate mitochondrial genome reference data for vertebrates, we (a) explore how reference specimen age and preservation method affect the amounts of mitochondrial DNA (as opposed to the amount of DNA originating from microorganisms) and

(b) estimate the cost of sequencing needed to generate mitochondrial genomes across taxa and tissue types.

2 | METHODS

2.1 | Sampling

In total, 210 vouchered vertebrate specimens collected across Denmark were included in this study (Figure 1). The specimens spanned 8 taxonomic classes, 48 orders, 99 families, 170 genera, and 199 species (Table S1). Most of the species (88%) belonged to three taxonomic groups: mammals ($n = 45$), birds ($n = 50$), and bony fish ($n = 80$).

Specimens were vouchered at the Natural History Museum of Denmark. The age and preservation conditions of the specimens varied from fresh material stored in ethanol at −18°C to ancient (historic) museum samples of bone or skin remain stored at ambient temperature. The details of each sample type are indicated in Table S1. For freshly collected specimens, a scalpel was used to cut out muscle biopsies and biopsies were stored in 96% ethanol at −18°C. For dried specimens, fur or bone material was scraped off with a scalpel and stored at −18°C. Pictures of voucher specimens were taken when possible. For all specimens, all relevant sampling information, such as sampling location, sampling date, and name of the person who carried out the taxonomic identification, was registered.

2.2 | Data generation

DNA was extracted from the 210 tissue samples using the Qiagen DNeasy Blood & Tissue Kit (version July 2016) with the following modifications: In the lysis step, 25 µl of Proteinase K was added to the lysis buffer and samples were incubated at 56°C overnight on a rotator. Negative extraction controls were included for each batch of extractions. All DNA extracts were fragmented on a Covaris LE220-plus system aiming at an average fragment length of 475 bp. A Qubit Fluorometer (Invitrogen) was used to quantify DNA in each extract. Preparation of sequencing libraries was carried out using the Blunt-End Single Tube (BEST) protocol (as described in Carøe et al., 2018; Mak et al., 2017) or the Blunt End Multi Tube (BEMT) protocol (described in Sirén et al., 2019)

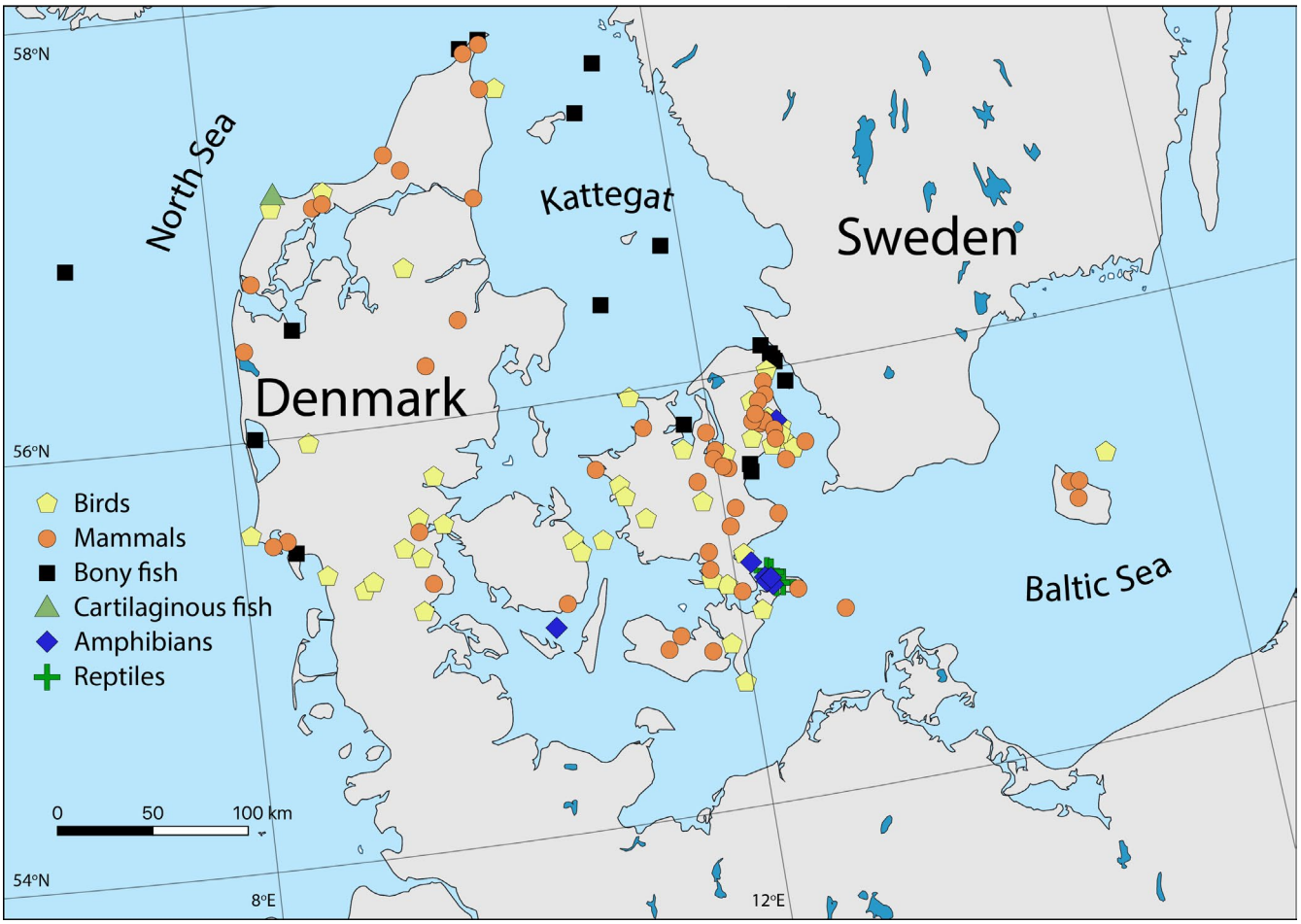

FIGURE 1 Samples were taken from 210 vouchered vertebrate specimens, representing 199 species, collected across Denmark. Sampling locations are shown for the 157 bird, mammal, fish, amphibian, and reptile specimens that had associated geo-coordinates

with double-indexing with matching indices to account for potential carryover between libraries on the flow cell (Kircher, Sawyer, & Meyer, 2012; Sinha, Stanley, Gulati, Ezran, & Travaglini, 2017). Libraries were pooled in equimolar concentrations aiming at ca. 5 Gb/library and principally sequenced on the Illumina HiSeq 4,000 platform using 150 bp paired-end chemistry either at the National High-throughput DNA Sequencing Centre (Copenhagen, Denmark) or Novogene (China). A small fraction of the samples (n = 19) was not sequenced using Illumina chemistry, instead, they were converted into libraries using BGISEQ-500-compatible adapters and sequenced on the BGISEQ platform using 100 bp PE chemistry at BGI Europe (Copenhagen, Denmark).

2.3 | Mitogenome assembly and annotation

Sequence reads were trimmed for adapters, consecutive stretches of Ns and low-quality bases using AdapterRemoval v2.2. Only sequences with a minimum length of 30 bp were retained. To increase the quality of the mtDNA assemblies, two different programs, Novoplasty v2.6.3 (Dierckxsens, Mardulyn, & Smits, 2017) and MitoZ v2.3 (Meng, Li, Yang, & Liu, 2019) were used for mitogenome assembly, both using default parameters. For the ones done with Novoplasty, we used a COI barcode sequence for each species retrieved from BOLD (https://boldsystems.org) as a seed (a starting sequence for assembly initiation). In the case of *Chirolophis ascanii*—DM356 where no barcode was available, we mapped the raw reads to barcodes of species from the closest taxonomic group with relaxed mapping parameters using Geneious v9.1.8 (geneious.com). Following all assemblies, we used Geneious for manual quality control. Furthermore, in cases where both Novoplasty and Mitoz had been used for mtDNA assembly, we used Geneious to compare the qualities of the assemblies. Annotations of the final assemblies were carried out using MITOS WebServer (mitos2.bioinf.uni-leipzig.de/index.py) and MitoZ v2.3 (Meng et al., 2019).

2.4 | Mapping statistics and phylogenetic analysis

To assess the fraction of mtDNA reads within each specimen, we mapped the raw adapter-free DNA reads (described above) to the assembled mtDNA sequences using the bwa-samtools pipeline. In short: we used the "bwa mem" algorithm in the bwa v0.7.10 (Li & Durbin, 2009) with stringent mapping parameters (-k19 -B20 -O16 -L5,5). To sort the mapped DNA reads and remove sequences with mapping quality of <30, we used samtools v1.3.1 (Li et al., 2009). Duplicates were removed with MarkDuplicates command in picard v2.20 (https://broadinstitute.github.io/picard). GATK v3.3.0 was used for realignment of the reads, which was followed by updating the md tags and calculating the extended BAQs with samtools. The soft clipped DNA reads in the bam files were removed based on the CIGAR field to avoid false positively mapped reads in the alignments.

The sequencing depth for each specimen was assessed as the median of the average depth values across 100 bp window sizes using Bedtools v2.28 (Quinlan & Hall, 2010).

The ANOVA and regression analyses were conducted in R (www.r-project.org) to assess the effects of age and storage conditions of the specimens on the amount of total vertebrate mtDNA. This assessment was performed by mapping the adapter-trimmed DNA reads against the deNovo assembled vertebrate mtDNA contigs for each sample. For these analyses, we only used samples for which the collection date was available (Table S1). The depth of coverage for the samples that we failed to assemble complete or partial mtDNA contigs was roughly assessed by mapping the reads to the barcode regions instead (Table S2). Since the raw number of reads between the samples varied considerably (ca. 2,000,000—226,000,000 reads), we used a weighting factor to account for differences in total number of sequences between the samples, which was calculated as a ratio of 2,024,134 (the number of reads of the least sequenced sample DM217) over the number of reads for each sample. These weighting factors were used to transform the median depth of coverage (DoC) values on mtDNA into the respective weighted estimates which was used as a dependent variable in the ANOVA analysis. Furthermore, we applied log-transformation to the weighted median DoC values in order to have equal variation among the groups, which was tested using Levene's test ("car" package in R) ($F_{5,97} = 2.396$, $p = .039$).

Phylogenetic reconstruction was carried out to assess the genetic relationship of the sequenced species within each of the vertebrate classes based on mtDNA as well as to identify potentially mislabeled species based on incorrect tree topologies. The protein-coding regions of the mtDNA were aligned with Mafft v7.309 (Katoh & Standley, 2013) and used as input for RAxML v8.2.12 (Stamatakis, 2014) with maximum likelihood approach with a GTR + GAMMA model of nucleotide substitution. Two hundred bootstrap replicates were performed to obtain node support.

3 | RESULTS

We generated ca. 3.49 billion DNA sequence reads from 210 specimens representing 199 Danish mammal, bird, fish, amphibian, and reptile species (Tables S1 and S2). We succeeded in assembling mitochondrial DNA (mtDNA) contigs for 192 specimens (representing 182 species) of which complete mitogenomes were created for 73 specimens. For 113 specimens, relatively long mtDNA contigs of >12 kb in size were created, while for the remaining six specimens, the length of mtDNA contigs were 3–10.5 kb largely due to poor preservation of the samples. The average value of the median depth of coverage for the 192 assemblies was 1,170.8×, ranging ca. 27–12,208×, while the average length was 16,177.5 bp (Table S2). The fraction of mtDNA reads (compared to the total number of retained reads after adapter removal) was around 0.54% ranging from ca. 0.005% to 5.62% with the highest fraction of mtDNA reads originating from tissue/muscle samples (Table S2).

Many of the Danish species included in the study did not have publicly available complete or partial mtDNA genomes. Among the 182 species for which we assembled mtDNA reference data (Figure 2), 89 species (with 30 complete and 59 partial mtDNA genomes) are presented here for the first time. Moreover, one of the 30 species (*Chirolophis ascanii*—DM356) for which we were able to assemble a complete mtDNA genome did not have a barcode available in the BOLD database (boldsystems.org/index.ph). Notably, four reference specimens among the mammals and fish were likely initially mislabeled or misidentified. This was discovered after blasting the respective COI barcodes against NCBI. Furthermore, for one of these samples, DM21, which was wrongly identified as *Eliomys quercinus*, this misidentification was also clearly revealed during the phylogenetic analyses. The museum representatives were informed and the specimens will be reidentified and their information corrected accordingly.

Maximum likelihood (ML) phylogenetic trees of six taxonomic classes based on mtDNA sequences are presented in Figure 3 and Figure S1.

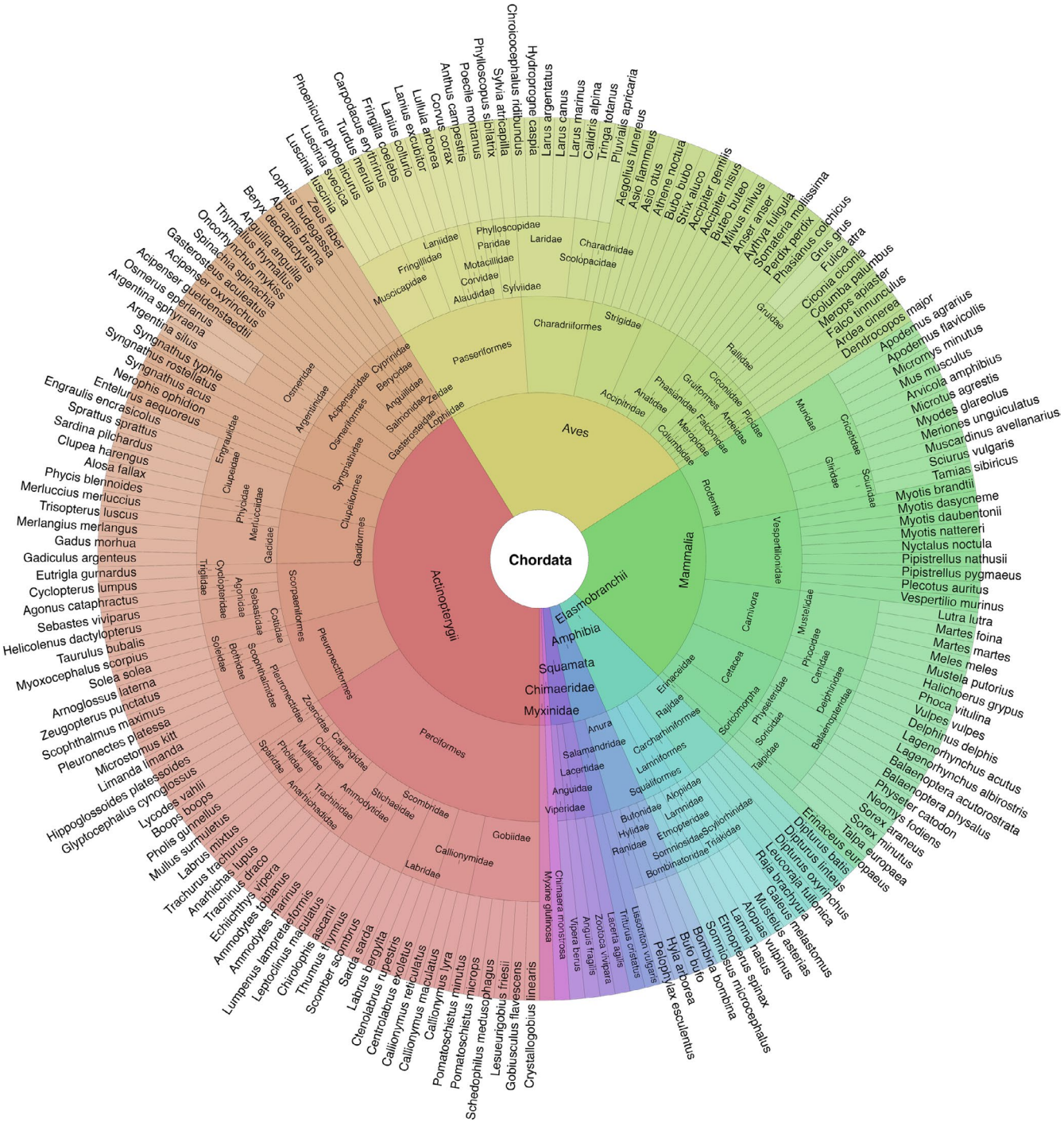

FIGURE 2 The 182 Danish vertebrate species (and associated phylum, class, order, and family) for which complete or partial mitogenome reference data was generated

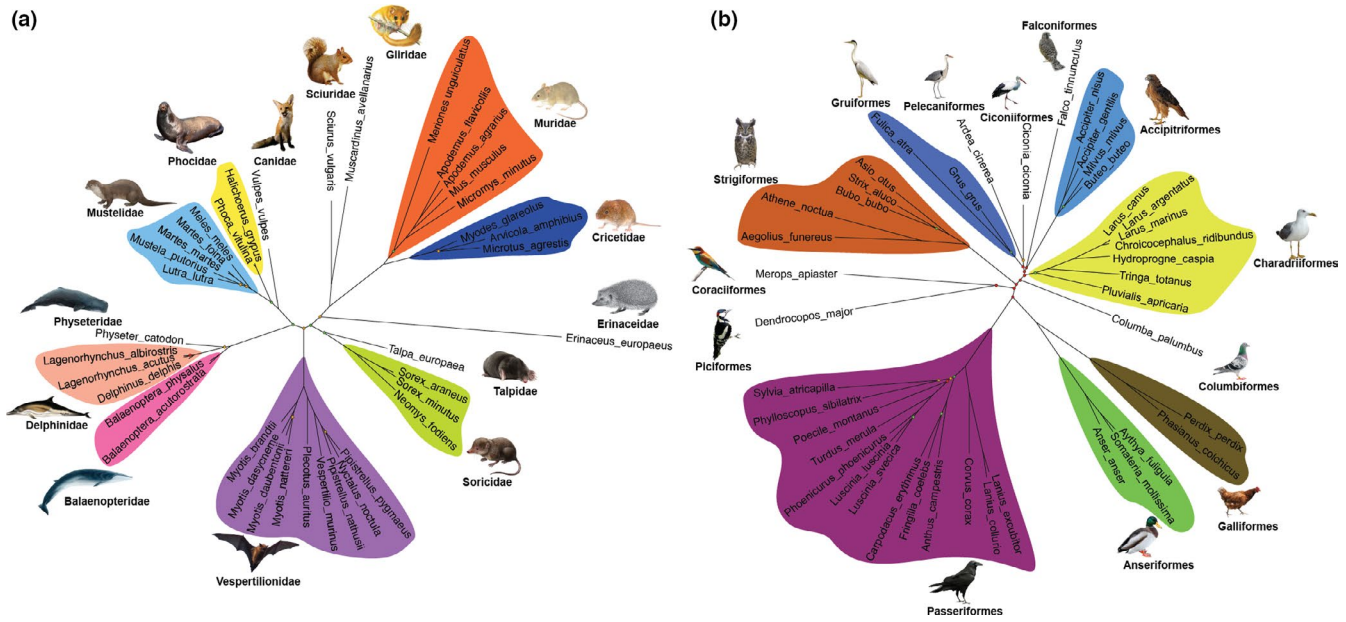

FIGURE 3 Phylogenetic trees based on the generated complete or partial mitochondrial reference genomes for mammalian (a) and avian (b) species. Maximum likelihood method implemented in RAxML was used with a GTR + GAMMA model of nucleotide substitution. Nodes have 100% bootstrap support based on 200 replicates, except the highlighted ones which have as follows: red, <50%; orange, 50%–90%, and green, >90% bootstrap support estimates

The ML trees of the reptiles, amphibians, and Elasmobranchii (subclass of cartilaginous fish) had 100% bootstrap support estimates for all the nodes and were identical with their respective species trees. In case of the mammalian, avian, and Actinopterygii (ray-finned fish) phylogenetic trees, the ML trees generally matched their respective species trees; however, the bootstrap estimates were <100% for many nodes, suggesting that the mtDNA alone may not suffice for resolving phylogenetic relationships at various taxonomic levels (Figure 3 and Figure S1).

We assessed the effects of storage conditions and age of the specimens on the amount of total vertebrate mtDNA. When using the storage conditions of the specimens as an independent variable, the overall ANOVA model was highly significant, indicating that indeed storage conditions affected DNA preservation (here assessed by the weighted median DoC) and accounted for ca. 50% of variation, $F(5,197) = 17.04$, $p < .001$, $\omega = 0.532$, with the reference speciemens that were stored dry performing the worst. Similar results were also obtained when using nonparametric Kruskal–Wallis Test (Chi square = 59.3, $p < .001$, $df = 5$).

We furthermore used regression analysis to test if the specimen's age significantly predicts the DNA preservation. The results showed that the age of the specimen only explains about 4% of the variance ($R^2 = 0.04$, $F_{1,201} = 9.35$, $p < .01$). However, this effect became nonsignificant when conducting a multiple regression with specimen's age and storage conditions as predictors.

4 | DISCUSSION

A comprehensive DNA reference database is of crucial importance in environmental DNA studies (Pawlowski et al., 2018; Schenekar, Schletterer, Lecaudey, & Weiss, 2020; Thomsen & Willerslev, 2015). As part of the project to establish the Danish national DNA reference database, DNAmark, we generated genome skimming data for vouchered specimens of 199 Danish birds, fish, mammals, amphibians and reptiles covering ca. 22% of the vertebrate species observed in Denmark. For 182 species, we were able to assemble complete or partial mitochondrial genomes. Around half of these species ($n = 89$) did not have published complete or partial mitochondrial reference genomes prior to this publication and hence through this project, we nearly doubled the public mitochondrial DNA data of complete or partial mitogenomes of vertebrate species observed in Denmark. Apart from the newly reported mtDNA sequences, we present mtDNA genomes from 93 vertebrate species that already had publicly available complete or partial mitochondrial genomes. This data can contribute to studies assessing mitochondrial intraspecies variation.

Notably, four mammal and fish reference specimens were likely initially mislabeled or misidentified at the museum. This was discovered after blasting their respective COI barcodes against NCBI and during the phylogenetic analyses.

Even though this study was not conducted as a controlled experiment for studying DNA preservation in various museum specimens, our relatively large dataset from various taxonomic groups allowed us to assess the overall effects of storage conditions and age of the reference specimens on the amount of total mitochondrial DNA. As expected, the storage conditions of the samples had an important role for DNA preservation $F(5,197) = 17.04$, $p < .001$, $\omega = 0.532$ with dried reference specimens having the least amount of mitochondrial DNA compared with any other storage condition, $p < .001$.

The age of the reference specimens in our dataset, however, had little effect on the amount of total vertebrate mtDNA. This was somewhat unexpected since in general the age of a sample is one of the major factors affecting DNA preservation of biological material (Allentoft et al., 2012; Bär, Kratzer, Mächler, & Schmid, 1988; Higgins, Rohrlach, Kaidonis, Townsend, & Austin, 2015; Itani, Yamamoto, Doi, & Miyaishi, 2011). The fact that age had little effect on the weighted median mtDNA coverage in our dataset indicates that other factors have more pronounced effect on the DNA preservation within relatively short time periods of a few decades. It has previously been shown that DNA degrades exponentially through time after the death of the organism (Allentoft et al., 2012; Bär et al., 1988; Higgins et al., 2015; Itani et al., 2011). This suggests that the preservation state of the specimen shortly after its death may have more detrimental outcome for DNA degradation than the storage conditions over the following longer periods. This may partially explain the little effect of specimen's age on DNA preservation measured based on weighted median mtDNA coveragein our dataset. This may also explain the few poorly preserved samples (e.g., DM239 and DM345), even though they were relatively freshly collected frozen muscle samples. Other factors such as the initial tissue-specific amount of mtDNA (Masuyama, Iida, Takatsuka, Yasuda, & Matsuki, 2005; Robin & Wong, 1988; Veltri, Espiritu, & Singh, 1990) or different DNA decay rates in varioustissue types (Itani et al., 2011) will undoubtedly be important factors to consider as well. However, since tissue type was highly correlated with sample storage conditions in our dataset (most skin samples were stored dried, Table S1), it was hard to assess the impact of tissue type on weighted median mtDNA coverage. It is also likely that more "exposed" tissues such as skin will have more extraneous DNA levels (and thus less endogenous) under the same storage conditions as perhaps other tissue types such as muscle or organ. Therefore, our results reflecting the effects of tissue type and storage conditions on DNA preservation should be interpreted with caution.

In order to obtain genome skimming data (ca. 5 Gb per specimen) within the frame of the DNAmark project, we pooled approximately 20 specimens per Illumina HiSeq lane. Such genome skimming approach has been shown to be effective in uncovering evolutionary histories of various taxonomic groups (Fonseca & Lohmann, 2020; Nauheimer et al., 2019; Nevill et al., 2020; Sarmashghi, Bohmann, Gilbert, Bafna, & Mirarab, 2019). However, there were many samples with high mtDNA coverage in our dataset. This indicated that for the well-preserved museum samples (such as e.g. ethanol preserved tissue/muscle) more samples per lane can be pooled for successful mtDNA assemblies. Hence, based on our results it is advisable to optimize this process further by pooling more samples per sequencing lane in future similar projects working with well-preserved samples. This approach, however, will not be feasible if genome skimming data (ca. 5 Gb per sample) are also desireable, as pooling more samples per lane will proportionally lower the total amount of sequencing for each sample. Given the average coverage of mtDNA of ca. 1,100×, it should be possible to pool, for example, 5 times as many samples (ca. 100) per lane (though reducing the amount of genome skimming data per sample) for obtaining roughly 200× mtDNA depth of coverage. This would ultimately lower the sequencing cost down to roughly 25 EUR per sample based on the prices as of 2019.

In the future, we hope to include all vertebrates from Denmark in the Danish DNA reference database, thereby creating the groundwork for DNA-based vertebrate monitoring in Denmark. Moreover, even though DNAmark project has a focus on various taxonomic groups of animals, plants, and fungi found in Denmark, most of the species are widely distributed across the temperate regions of the globe as well as in the waters of the North Atlantic. Therefore, this database is not only an important resource for vertebrate monitoring projects in Denmark, but also far beyond its borders. In addition, combining similar regional DNA reference databases (e.g., Alsos et al., 2020; Mohd Salleh et al., 2017) will fill in the gaps in the sequenced species around the globe.

ACKNOWLEDGMENTS

We are grateful for the support from Aage V. Jensen Naturfond for the establishment of the national DNA reference database, DNAmark. We thank the Natural History Museum of Denmark for curation of newly collected voucher specimens, for access to collections and for assistance with regards to sample collection. Specifically, we thank Peter Rask Møller, Marcus Anders Krag, Henrik Carl, Kasper Thorup, Jan Bolding Kristensen, Daniel Klingberg Johansson, Eline Lorenzen, Morten Tange Olsen, and Morten Erik Allentoft. Further, we are grateful to the DNAmark committee for their valuable inputs and for collection and taxonomic identification of voucher specimens. Finally, we thank the Danish National High-throughput DNA Sequencing Center for sequencing and fruitful discussions. We also thank the anonymous reviewers for their evaluation and constructive comments.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

Raw sequencing data are deposited in the SRA under project number PRJNA607895. Assembled mitochondrial sequences are deposited in GenBank under accession numbers indicated in the Table S2.

ORCID

Ashot Margaryan https://orcid.org/0000-0002-2576-2429
Christina Lehmkuhl Noer https://orcid.org/0000-0001-8303-5759
Stine Raith Richter https://orcid.org/0000-0002-3556-5146
Shyam Gopalakrishnan https://orcid.org/0000-0002-2004-6810
Christian Carøe https://orcid.org/0000-0001-9601-6768
M. Thomas P. Gilbert https://orcid.org/0000-0002-5805-7195
Kristine Bohmann https://orcid.org/0000-0001-7907-064X

REFERENCES

Agersnap, S., Larsen, W. B., Knudsen, S. W., Strand, D., Thomsen, P. F., Hesselsøe, M., … Møller, P. R. (2017). Monitoring of noble, signal and narrow-clawed crayfish using environmental DNA from freshwater samples. *PLoS One*, *12*(6), e0179261. https://doi.org/10.1371/journal.pone.0179261

Alberdi, A., Aizpurua, O., Bohmann, K., Gopalakrishnan, S., Lynggaard, C., Nielsen, M., & Gilbert, M. T. P. (2018). Promises and pitfalls of using high-throughput sequencing for diet analysis. *Molecular Ecology Resources*, *19*(2), 327–348. https://doi.org/10.1111/1755-0998.12960

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., … Bunce, M. (2012). The half-life of DNA in bone: Measuring decay kinetics in 158 dated fossils. *Proceedings of the Biological Sciences/the Royal Society*, *279*(1748), 4724–4733.

Alsos, I. G., Lavergne, S., Merkel, M. K. F., Boleda, M., Lammers, Y., Alberti, A., … Coissac, E. (2020). The treasure vault can be opened: Large-scale genome skimming works well using herbarium and silica gel dried material. *Plants*, *9*(4), 432. https://doi.org/10.3390/plants9040432

Bär, W., Kratzer, A., Mächler, M., & Schmid, W. (1988). Postmortem stability of DNA. *Forensic Science International*, *39*(1), 59–70. https://doi.org/10.1016/0379-0738(88)90118-1

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., & Sayers, E. W. (2018). GenBank. *Nucleic Acids Research*, *46*(D1), D41–D47. https://doi.org/10.1093/nar/gkx1094

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

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., … Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution/British Ecological Society*, *9*(2), 410–419.

Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects: In silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, *14*(6), 1160–1170.

Coissac, E., Hollingsworth, P. M., Lavergne, S., & Taberlet, P. (2016). From barcodes to genomes: Extending the concept of DNA barcoding. *Molecular Ecology*, *25*(7), 1423–1428. https://doi.org/10.1111/mec.13549

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

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, *45*(4), e18.

Fietz, K., Graves, J. A., & Olsen, M. T. (2013). Control control control: A reassessment and comparison of GenBank and chromatogram mtDNA sequence variation in Baltic grey seals (*Halichoerus grypus*). *PLoS One*, *8*(8), e72853. https://doi.org/10.1371/journal.pone.0072853

Fonseca, L. H. M., & Lohmann, L. G. (2020). Exploring the potential of nuclear and mitochondrial sequencing data generated through genome-skimming for plant phylogenetics: A case study from a clade of neotropical lianas. *Journal of Systematics and Evolution*, *58*(1), 18–32. https://doi.org/10.1111/jse.12533

Foote, A. D., Thomsen, P. F., Sveegaard, S., Wahlberg, M., Kielgast, J., Kyhn, L. A., … Gilbert, M. T. P. (2012). Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS One*, *7*(8), e41781. https://doi.org/10.1371/journal.pone.0041781

Hebert, P. D. N., Ratnasingham, S., & deWaard, J. R. (2003). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *270*(Suppl 1), S96–S99.

Higgins, D., Rohrlach, A. B., Kaidonis, J., Townsend, G., & Austin, J. J. (2015). Differential nuclear and mitochondrial DNA preservation in post-mortem teeth with implications for forensic and ancient DNA studies. *PLoS One*, *10*(5), e0126935. https://doi.org/10.1371/journal.pone.0126935

Itani, M., Yamamoto, Y., Doi, Y., & Miyaishi, S. (2011). Quantitative analysis of DNA degradation in the dead body. *Acta Medicinae Okayama*, *65*(5), 299–306.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1), e3. https://doi.org/10.1093/nar/gkr771

Leray, M., Knowlton, N., Ho, S.-L., Nguyen, B. N., & Machida, R. J. (2019). GenBank is a reliable resource for 21st century biodiversity research. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(45), 22651–22656. https://doi.org/10.1073/pnas.1911714116

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Mak, S. S. T., Gopalakrishnan, S., Carøe, C., Geng, C., Liu, S., Sinding, M.-H.-S., … Gilbert, M. T. P. (2017). Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *GigaScience*, *6*(8), 1–13. https://doi.org/10.1093/gigascience/gix049

Masuyama, M., Iida, R., Takatsuka, H., Yasuda, T., & Matsuki, T. (2005). Quantitative change in mitochondrial DNA content in various mouse tissues during aging. *Biochimica Et Biophysica Acta*, *1723*(1–3), 302–308. https://doi.org/10.1016/j.bbagen.2005.03.001

Meiklejohn, K. A., Damaso, N., & Robertson, J. M. (2019). Assessment of BOLD and GenBank–Their accuracy and reliability for the identification of biological materials. *PLoS One*, *14*(6), e0217084. https://doi.org/10.1371/journal.pone.0217084

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

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., … Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. *Royal Society Open Science*, *2*(7), 150088. https://doi.org/10.1098/rsos.150088

Mohd Salleh, F., Ramos-Madrigal, J., Peñaloza, F., Liu, S., Mikkel-Holger, S. S., Riddhi, P. P., … Gilbert, M. T. P. (2017). An expanded mammal mitogenome dataset from Southeast Asia. *GigaScience*, *6*(8), 1–8.

Nauheimer, L., Cui, L., Clarke, C., Crayn, D. M., Bourke, G., & Nargar, K. (2019). Genome skimming provides well resolved plastid and nuclear phylogenies, showing patterns of deep reticulate evolution in the tropical carnivorous plant genus *Nepenthes* (Caryophyllales). *Australian Systematic Botany*, *32*(2–3), 243–254. https://doi.org/10.1071/SB18057

Nevill, P. G., Zhong, X., Tonti-Filippini, J., Byrne, M., Hislop, M., Thiele, K., … Small, I. (2020). Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics. *Plant Methods*, *16*, 1. https://doi.org/10.1186/s13007-019-0534-5

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., … Kahlert, M. (2018). The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of the Total Environment*, *637–638*, 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ratnasingham, S., & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*(3), 355–364.

Robin, E. D., & Wong, R. (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of Cellular Physiology*, *136*(3), 507–513. https://doi.org/10.1002/jcp.1041360316

Sarmashghi, S., Bohmann, K., Gilbert, M. T. P., Bafna, V., & Mirarab, S. (2019). Skmer: Assembly-free and alignment-free sample identification using genome skims. *Genome Biology*, *20*(1), 34. https://doi.org/10.1186/s13059-019-1632-4

Schenekar, T., Schletterer, M., Lecaudey, L. A., & Weiss, S. J. (2020). Reference databases, primer choice, and assay sensitivity for environmental metabarcoding: Lessons learnt from a re-evaluation of an eDNA fish assessment in the Volga headwaters. *River Research and Applications*, *36*, 1004–1013. https://doi.org/10.1002/rra.3610

Sigsgaard, E. E., Nielsen, I. B., Carl, H., Krag, M. A., Knudsen, S. W., Xing, Y., … Thomsen, P. F. (2017). Seawater environmental DNA reflects seasonality of a coastal fish community. *Marine Biology*, *164*(6), 128. https://doi.org/10.1007/s00227-017-3147-4

Sinha, R., Stanley, G., Gulati, G. S., Ezran, C., & Travaglini, K. J. (2017). Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. *BioRxiv*, https://www.biorxiv.org/content/biorxiv/early/2017/04/09/125724.full.pdf

Sirén, K., Mak, S. S. T., Melkonian, C., Carøe, C., Swiegers, J. H., Molenaar, D., … Gilbert, M. T. P. (2019). Taxonomic and functional characterization of the microbial community during spontaneous in vitro fermentation of riesling must. *Frontiers in Microbiology*, *10*, 697. https://doi.org/10.3389/fmicb.2019.00697

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

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, *21*(8), 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

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

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

Thomsen, P. F., & Sigsgaard, E. E. (2019). Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecology and Evolution*, *9*(4), 1665–1679. https://doi.org/10.1002/ece3.4809

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Veltri, K. L., Espiritu, M., & Singh, G. (1990). Distinct genomic copy number in mitochondria of different mammalian organs. *Journal of Cellular Physiology*, *143*(1), 160–164. https://doi.org/10.1002/jcp.1041430122

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Margaryan A, Noer CL, Richter SR, et al. Mitochondrial genomes of Danish vertebrate species generated for the national DNA reference database, DNAmark. *Environmental DNA*. 2021;3:472–480. https://doi.org/10.1002/edn3.138