Affordable de novo generation of fish mitogenomes using amplification-free enrichment of mitochondrial DNA and deep sequencing of long fragments

Ana Ramon-Laca[1], Ramon Gallego[2], and Krista Nichols[2]

[1]CICOES
[2]NOAA Fisheries Northwest Fisheries Science Center

June 9, 2022

Abstract

Biomonitoring surveys from environmental DNA make use of metabarcoding tools to describe the community composition. These studies match their sequencing results against public genomic databases to identify the species. However, mitochondrial genomic reference data are yet incomplete, only a few genes may be available, or the suitability of existing sequence data is suboptimal for species level resolution. Here we present a dedicated and cost-effective workflow with no DNA amplification for generating complete fish mitogenomes for the purpose of strengthening fish mitochondrial databases. Two different long-fragment sequencing approaches using Oxford Nanopore sequencing coupled with mitochondrial DNA enrichment were used. One where the enrichment is achieved by preferential isolation of mitochondria followed by DNA extraction and nuclear DNA depletion ('mitoenrichment'). A second enrichment approach takes advantage of the CRISPR-Cas9 targeted scission on previously dephosphorylated DNA ('targeted mitosequencing'). The sequencing results varied between tissue, species, and integrity of the DNA. The mitoenrichment method yielded 0.17-12.33 % of sequences on target and a mean coverage ranging from 74.9-805-fold. The targeted mitosequencing experiment from native genomic DNA yielded 1.83-55 % of sequences on target and a 38-2123-fold mean coverage. This helped complete the mitogenome of species with homopolymeric regions, tandem repeats and gene rearrangements. We demonstrate that deep sequencing of long fragments of native fish DNA is possible, can be achieved with low computational resources in a cost-effective manner, exceeding the widespread genome skimming approach, and allowing the discovery of mitogenomes of non-model or understudied fish taxa to a broad range of laboratories worldwide.

Affordable de novo generation of fish mitogenomes using amplification-free enrichment of mitochondrial DNA and deep sequencing of long fragments

Running title: Amplification-free sequencing of fish mitogenomes

Ana Ramón-Laca[1], Ramón Gallego[2] and Krista M. Nichols[2]

[1]CICOES, University of Washington and Northwest Fisheries Science Center, National Marine Fisheries Service, Seattle, WA, U.S.A.

[2]Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA 98112, U.S.A.

Corresponding author e-mail address: analaca@mncn.csic.es

Ana Ramón-Laca current affiliation:

Science and Business S.L. Calle Perú 6, 28290, Las Matas-Las Rozas, Spain

Museo Nacional de Ciencias Naturales (MNCN), CSIC, C/ José Gutiérrez Abascal 2, 28006, Madrid, Spain.

Ramón Gallego current affiliation:

Departamento de Biología, Facultad de Ciencias, Universidad Autónoma de Madrid, C/ Darwin, 2, 28049 Cantoblanco, Madrid, Spain

Abstract

Biomonitoring surveys from environmental DNA make use of metabarcoding tools to describe the community composition. These studies match their sequencing results against public genomic databases to identify the species. However, mitochondrial genomic reference data are yet incomplete, only a few genes may be available, or the suitability of existing sequence data is suboptimal for species level resolution. Here we present a dedicated and cost-effective workflow with no DNA amplification for generating complete fish mitogenomes for the purpose of strengthening fish mitochondrial databases. Two different long-fragment sequencing approaches using Oxford Nanopore sequencing coupled with mitochondrial DNA enrichment were used. One where the enrichment is achieved by preferential isolation of mitochondria followed by DNA extraction and nuclear DNA depletion ('mitoenrichment'). A second enrichment approach takes advantage of the CRISPR-Cas9 targeted scission on previously dephosphorylated DNA ('targeted mitosequencing'). The sequencing results varied between tissue, species, and integrity of the DNA. The mitoenrichment method yielded 0.17-12.33 % of sequences on target and a mean coverage ranging from 74.9 to 805-fold. The targeted mitosequencing experiment from native genomic DNA yielded 1.83-55 % of sequences on target and a 38 to 2123-fold mean coverage. This produced complete the mitogenome of species with homopolymeric regions, tandem repeats, and gene rearrangements. We demonstrate that deep sequencing of long fragments of native fish DNA is possible and can be achieved with low computational resources in a cost-effective manner, opening the discovery of mitogenomes of non-model or understudied fish taxa to a broad range of laboratories worldwide.

Resumen

Los estudios de biomonitoreo mediante ADN ambiental utilizan herramientas de caracterización genética (metabarcoding) para describir la composición de la comunidad. Estos estudios contrastan las secuencias obtenidas con bases de datos genómicas públicas para así identificar la especie. Sin embargo, las bases de datos mitocondriales de referencia distan mucho de estar completas. En la mayor parte de los casos solo hay unos pocos genes disponibles o los datos existentes no ofrecen resolución hasta el nivel de especie. En este estudio presentamos un método dedicado a generar mitogenomas de peces completos de forma rentable y sin necesidad de amplificación del ADN, con el objeto de ampliar las bases de datos mitocondriales de peces. Para ello se utilizaron dos enfoques diferentes de secuenciación de fragmentos largos utilizando secuenciación Oxford Nanopore y enriquecimiento de ADN mitocondrial. Uno en el que el enriquecimiento se logra mediante el aislamiento preferencial de mitocondrias seguido de extracción del ADN y la eliminación del ADN nuclear ("mitoenriquecimiento"). En el segundo enfoque se aprovecha la capacidad de escisión dirigida por la endonucleasa CRISPR-Cas9 sobre ADN previamente desfosforilado ("mitosecuenciación dirigida"). Los resultados difirieron con el tejido, la especie y la integridad del ADN. El método de mitoenriquecimiento produjo un 0,17-12,33% de secuencias objetivo y una cobertura media entre 74,9 y 805 secuencias. El experimento de mitosecuenciación dirigida a partir de ADN genómico nativo produjo entre 1,83 y 55 % de secuencias objetivo y una cobertura media de 38 a 2123 secuencias. Este estudio permitió completar mitogenomas de diferentes especies que incluyen regiones homopoliméricas, repeticiones en tándem y reorganización de genes. Demostramos que la secuenciación intensiva de fragmentos largos de ADN de peces es posible, se puede lograr con bajos recursos informáticos de una manera económca, superando el método generalizado de secuenciación genómica de baja cobertura y permitiendo el descubrimiento de mitogenomas de taxones de peces no modelo o poco estudiados a una amplia gama de laboratorios en todo el mundo.

Keywords: mitogenome, fish, eDNA, long fragment, Cas9 targeted sequencing, mitochondrial enrichment

Introduction

Environmental DNA (eDNA) is providing previously unthinkable insights into the aquatic environment as

a non-invasive and relatively cost-efficient tool, illustrating the presence-absence and distribution of certain species and the composition of the community of an ecosystem, and is particularly informative to evaluate these parameters as a result of variable conditions (Furlan et al., 2019; Gallego et al., 2020; Stat et al., 2017; Thomsen & Willerslev, 2015). Besides the logistics and technical constraints to acquire samples, there are further challenges in accurately characterizing the biome, including the molecular strategy used (i.e. DNA extraction and the marker or gene of choice), and the reference databases used to identify the origin of the DNA found in a certain environment (Jackman et al., 2021; Schenekar et al., 2020; Wang et al., 2021). Environmental DNA identification needs robust, comprehensive, and accurate DNA reference libraries based on solid taxonomic frameworks, and this effort requires more exhaustive, comprehensive, and revamped efforts in light of recent technical advancements in DNA sequencing (Margaryan et al., 2021; Novak et al., 2020; Taberlet et al., 2012).

There are currently two main strategies for biomonitoring surveys to describe the community composition or evaluate the abundance of certain species: DNA metabarcoding approaches or targeted molecular assays using quantitative Polymerase Chain Reaction (qPCR) or digital PCR (dPCR) (Shu et al., 2020). These molecular tools are well established and have been added to the toolbox of conservation management. These studies make use of the public genomic databases and, more specifically, of mitochondrial genes found in public repositories to construct well represented alignments to identify the amplicon sequence variants (ASV) in the case of metabarcoding or achieve the desired specificity for primer design in the case of targeted assays. However, in spite of the colossal sequencing effort undertaken in the last two decades with initiatives like the Barcode of Life (Mugnai et al., 2021; Ratnasingham & Hebert, 2007), mitochondrial genome reference data for the breadth of taxa of interest are yet not accessible and if they are, only one or a few genes may be available. Among these, the suitability of existing sequence data for the purpose of designing species-specific oligonucleotides is typically suboptimal for targeted eDNA assays. Such is the situation of one the most traditionally sequenced genes for barcoding purposes, the cytochrome c oxidase subunit 1 (COI) of the mitochondrion that offers scattered, short conserved regions that are unsuitable for primer design in order to effectively discriminate the taxon of interest (Langlois et al., 2020; Margaryan et al., 2021; Schroeter et al., 2020). Metabarcoding analysis of fish targets more conserved genes or regions where universal primers can be placed flanking interspecifically variable regions. Examples are the 12S rRNA subunit (12S) using the MiFish primers (Miya et al., 2015) that produce an amplicon of ca. 170 base pairs (bp) and the 16S rRNA (16S) with primers Ac16S that amplify a region of 330 bp (Evans et al., 2016) and ca. 100 bp-fragment with Fish16S primers (Deagle et al., 2009; Shaw et al., 2016). The aforementioned markers have provided an extraordinary wealth of information for community studies and species detection using eDNA (Miya, 2022; Shu et al., 2020). Building more comprehensive mitochondrial genome databases would be particularly advantageous for those species for whom identification cannot be resolved with short fragments of the 16S, 12S, or COI genes and to expand the representation to account for intra and interspecific variability. In addition, having more regions of the mitogenome available would facilitate an eDNA multi-marker metabarcoding approach, eDNA population genetics studies and even explore the possibility of mitochondria-associated disorders caused by mutations, an unexplored line of research that may provide insights into health or fitness questions (Brown, 2008; Dugal et al., 2021; Jackman et al., 2021; Jensen et al., 2020; Sharma & Sampath, 2019). Despite the mitochondrial genes being physically linked to each other, mitochondrial haplotypes from eDNA could determine minimum number of individuals and provide information about the origin of the populations (e.g. in the case of the stocks of anadromous fishes found in the ocean (Weitemier et al., 2021)). Additionally, whole mitochondrial genome databases, from verified, vouchered specimens, will also be critical for seafood monitoring, as molecular methods are routinely used to identify species in domestic and international trade (Bourret et al., 2020; Ogden, 2008). The expansion of mitochondrial genomic databases is not only costly and requires access to voucher specimens, but also relies in the use of so-called universal primers that have proven to be less generic than desired. Whole mitochondrial sequencing, although attainable, is still expensive and not readily accessible for all research groups due to a limitation in read length, non-affordable methods and long protocols (Gilpatrick et al., 2020).

In this study, we explore target enrichment methods to attain whole mitogenome sequence data in a simple

and cost-effective manner. Current technologies allow for whole genome direct sequencing (i.e. no special treatments are necessary) that can yield regions of interest, particularly those in high copy number such as mitochondrial DNA, which can be identified and recovered from the data in a process called 'genome skimming' or shallow sequencing (Straub et al., 2012). We propose that targeted mitochondrial DNA enrichment during the DNA isolation process or library preparation step should be sought if willing to reduce costs, time, data storage capacity and bioinformatic capabilities while improving coverage and consensus sequences, avoiding pseudogenes and general background noise that can affect the molarity of the target and thus compromise the sequencing performance. In spite of hundreds to about a thousand mitochondrial DNA (mtDNA) copies present in a fish cell depending on the tissue and the age, which is lower than in mammals (Hartmann et al., 2011), the amount of mtDNA in a preparation is normally around 0.1% of the genomic DNA (Robin & Wong, 1988) and is overwhelmed with nuclear DNA (nDNA). Different enrichment outcomes can be attained depending on the DNA extraction, treatment, sequencing and bioinformatic approaches employed. The physical properties of mtDNA (enclosed organelle physical location and the circularity of the mitochondrial genome) can be used to preferentially extract mtDNA using sequential precipitation methods or to deplete the non-circular DNA (i.e., nDNA) using exonucleases. Targeted enrichment can be also conducted using CRISPR-based methods by targeting conserved regions of the mitogenomes with specific guide RNAs (Schultzhaus et al., 2021). Mitochondrial enrichment without PCR amplification avoids universal primer incompatibility and PCR amplification errors. Additionally, long range amplification is proving challenging (Ramón-Laca personal observations, (Gilpatrick et al., 2020)) and target enrichment using hybridization capture is not yet fully operative for long fragments.

Long fragment sequences can help diminish the number of nuclear mitochondrial sequences (NumtS) that can be very long and are found in fish in a greater ratio than in most vertebrate species (Antunes & Ramos, 2005; Dayama et al., 2014). Long fragments preserve the order of the genes, in contrast with the short reads sequencing platforms that can also be affected by PCR bias on AT-rich regions (Gan et al., 2019). Long sequences can be key for accurate genome assembly (Pollard et al., 2018), in particular in repetitive regions, which are sometimes found in the control region of the mitochondria and have proven challenging to sequence with traditional methods (i.e., Sanger and whole genome sequencing of short fragments) (Formenti et al., 2021; McDonald et al., 2021). A rearrangement in the order of the genes will not be missed if transferring annotations from a different species because the order of the genes is determined by the sequence and not the reference genome. For all the aforementioned reasons, a de novo assembly of whole mitogenomes should be favored, to not bias the newly generated mitogenomes and to not overlook any possible modifications.

However, long-fragment sequencing comes with its own challenges. The main downside, and a common criticism, of long-read sequencing with Oxford Nanopore technologies is the relatively high error rates in the sequences obtained. Nonetheless, in opposition with short-read platforms, these errors are mostly random except for the homopolymeric regions from single pore reads with ONT that can be overcome with high read depth (Pollard et al., 2018; Schultzhaus et al., 2021) and with the constantly evolving flow cells, chemistry and base calling algorithms. Collections of fish have traditionally preserved the specimens (e.g. whole individuals, fin clips) in jars or tubes with >95% ethanol. This method has worked well for gene sequencing or microsatellite or SNP typing, but does not prevent the degradation of the high molecular weight fragments (Oosting et al., 2020) and thus most specimens of the collections will not yield fragments of the desired length to hinder accidental sequencing of NumtS. In addition, fish samples can sometimes take long to be sorted even on board of dedicated research vessels, which can compromise tissue quality and lead to degradation of most of the high molecular weight DNA (Oosting et al., 2020; Rodriguez-Ezpeleta et al., 2013).

In this study we show how to generate whole mitogenomes from fish species, with the aim of generating affordable and comprehensive databases that are not restricted to a few genes of interest. The long-fragment approach combined with the mitochondrial DNA enrichment produced whole mitogenomes with full coverage and great sequencing depth while using fewer computational and sequencing resources than genome skimming. Two approaches that enrich the mitochondrial DNA were evaluated in this study: 1) Mitochondrial DNA enrichment by isolating intact mitochondria; and 2) Targeted mitosequencing by using CRISPR Cas9

on conserved regions of the mitochondrial genome. Both approaches are followed by sequencing on an Oxford Nanopore platform. These target enrichment and long-fragment sequencing approaches efficiently produce data for whole mitogenomes while using less computational and sequencing resources than genome skimming, simplifying the discovery of mitogenomes of non-model or understudied fish taxa to a broad range of laboratories worldwide.

Methods

DNA extraction methods

Two different DNA extraction methods were used: 1) an organic DNA extraction using a phenol-chloroform based method that avoids pipetting to maximize the integrity of the DNA (Ana Ramón-Laca et al., 2021) and favors the isolation of DNA of high molecular weight; 2) and a mitochondria isolation procedure (see below, Fig 1 and S1) followed by the phenol-chloroform extraction. A chinook salmon (*Oncorhynchus tshawytscha*) from the NOAA Northwest Fisheries Science Center hatchery was euthanized and approximately 0.125 cm^3 of skeletal muscle (fresh) and liver and heart (frozen at -80^{o}C) samples were dissected (Table 2). These tissue samples in Longmire buffer were macerated using a TissueRuptor (Qiagen) with disposable probes. DNA was eluted in 100 and 50 µl of TlowE for either extraction method, respectively. Additional methods discarded for the purposes of this study are in File S3. Additional species (Table 2) were also extracted with either one or the two methods and these were eluted in 200 and 100 µl of TlowE, respectively. Wide bore tips were used where required to avoid fragmentation. All DNA extractions were performed in a dedicated genetics laboratory.

DNA quantity and quality

All samples were quantified with either the dsDNA high sensitivity or broad range kits on the Qubit (Invitrogen, ThermoFisher). In the case of the chinook salmon, a TaqMan qPCR specific assay was used (Shelton et al., 2019) to quantify the amount of target DNA of each DNA extraction using gBlocks^{TM} fragments (synthetic double stranded oligonucleotides from Integrated DNA Technologies Inc.) as standards from 10^5 to 1 copies μl^{-1} and assess the target to total ratio. The integrity of the DNA (DIN) was also evaluated using the Genomic DNA ScreenTape and kit on a 2200 TapeStation System (Agilent Technologies)

Enrichment by organelle isolation and nuclear DNA depletion: 'mitoenrichment'

The mitochondria isolation procedure is based on sequential precipitation. We used the Mitochondria Isolation kit for cultured cells (ThermoFisher) and followed protocol B (details in S1). Subsequently, we added 2 ml of mitochondria lysis buffer (Isokallio & Stewart, 2018) and 100 µl of proteinase k to the crude mitochondria pellet and incubated up to 2 h after which the phenol chloroform DNA extraction was conducted. In order to maximize the amount of pure mtDNA, nuclear DNA (nDNA) removal was performed in a 40 µl reaction with 30 µl of DNA at ideally >60 ng μl^{-1}, 1×NEBuffer4, 1mM ATP and 10 U of Exonuclease V (RecBCD) (New England Biolabs Inc.). The digestion was performed at 37°C for 2 h followed by heat-inactivation at 70°C for 30 min (Dhorne-Pollet et al., 2020). The resulting DNA was cleaned and concentrated using 0.7× Ampure XP (Agencourt) (with two washes with 70% EtOH) and eluted in 10 µl of TlowE. Nanopore library preparation was conducted with the rapid sequencing kit (SQK-RAD004, ONT), which uses a transposase that cleaves the DNA randomly linearizing the mitochondrial DNA while adding the necessary adapters for the motor proteins that serve as sequencing adapters using up to 400 ng oftemplate following manufacturer's protocol (Fig 1, detailed protocol in S1). Running times varied as in Table 3.

Targeted mitosequencing of genomic DNA using CRISPR Cas9: 'targeted mitosequencing'

Design of Cas9 guide RNAs: The CRISPR Cas9 enzyme cleaves at specific sites of your choice with the aid of the so-called guide RNA (gRNA) sequences. The latter were designed based on an alignment of several salmonid species and rockfish species that were of interest and sequences compatible to all the species were found using the 'Find CRISPR Site' tool within Geneious Prime 2021.2.2 (Biomatters Ltd, www.geneious.com). This tool searches for sequences of 20 nucleotide followed by a protospacer adjacent motif (PAM) on the 3' end of three nucleotides with the motif NGG. Five candidate gRNAs (Table1, S1) were

selected to ensure there was going to be enough coverage in the event of one or two failing due to mismatches as well as to ensure even coverage (López-Girona et al., 2020). These cutting sites are conserved regions of the mitogenome in fishes.: one (alias 41) on the 12SrRNA gene (12S), two (alias 28 and 17) about 2 kbp further on the 16SrRNA (16S) and two more are overlapping and cleave on the tRNA-Gly (G) that is located between the COIII and ND3 coding genes (alias 20 and 24) about 7 kbp away. The cutting sites on the tRNA-Gly were chosen to increase the coverage at approximately halfway on the mitogenome from the 16S or 12S cuts to ensure the (cytochrome b) cytb and (control region) D-loop areas were covered evenly in which seemed a promising site although there was no guarantee of its match with most of the fish due to lack of representation of this gene in GenBank. The complete mitogenome of 2941 specimens was downloaded and the specificity of the gRNAs was tested in silico against the consensus of the alignment of 176 full Chondrichthyes mitogenomes (with 0,0, 0, 5, 5 mismatches for gRNA alias 41, 28, 17, 20 and 24, respectively), 13 Cyclostomata (2, 2, 0, 5, 5), 7 Sarcopterygii (0, 0, 0, 3, 3) and 2693 Teleostei (0, 0, 0, 2,2). These five gRNA were also tested against the whole genome of *Sebastes schlegelii*(GenBank Acc ASM1467356) to assess whether the guides could target nuclear regions using a Bowtie2 alignment within Geneious, with low sensitivity/fast parameters and no contigs were found The gRNA sequences were synthesized with a crRNA tail that will bind to the transactivating RNA (tracRNA) to form the duplex that the Cas9 uses to find the cutting sites (IDT, Skokie, IL, USA).

Table 1. Details of the guide RNA (gRNA) sequences used to direct the CRISPR-Cas9 scission. The gRNA positions (5'-3') are based on the mitogenome sequence of *Sebastes aleutianus* (GenBank Acc. NC039779).

Name	alias	gene location	direction	Position 5'-3'	Sequence	PAM	Leng
Sa-12S-591R	**41**	12S rRNA	reverse	591-572	AAGTCCTTTGGGTTTTAAGC	NGG	20
Sa-16S-2601F	**17**	16S rRNA	forward	2601-2620	TACGTGATCTGAGTTCAGAC	NGG	20
Sa-16S-2550R	**28**	16S rRNA	reverse	2550-2531	GATGTCCTGATCCAACATCG	NGG	20
Sa-Gly-9640R	**24**	tRNA-Gly	reverse	9640-9621	TTTAACCAAGACCGGGTGAT	NGG	20
Sa-Gly-9607F	**20**	tRNA-Gly	forward	9607-9626	ATAAGTGACTTCCAATCACC	NGG	20

Cas9 enrichment for mtDNA: The genomic DNA samples were treated to maximize the available mitochondrial target fragments for the sequencing process using the Cas9 targeted sequencing method described by Gilpatrick et al (2020) with a few variations (Fig 1). More specifically, the crRNA-tracrRNA duplexes were first assembled individually by incubating 1 µl of each 100 µM crRNA, 1 µl of 100 µM Alt-R® CRISPR-Cas9 tracrRNA and 8 µl of duplex buffer (IDT) at 95 °C for 5 min in separate tubes. Five Alt-R CRISPR-Cas9 tracrRNA and guide RNA molecules hybridizations were subsequently combined at 10 µM each to form the ribonucleoprotein (RNP) complexes by adding 4 µl of Alt-R HiFi Cas9 V3 Nuclease (0.8 µl per guide included, IDT) and 10 µl of 10 × CutSmart$^{(r)}$ buffer (New England Biolabs, Ipswich, MA, USA - NEB), 36 µl of nuclease-free H_2O and incubated at room temperature for 30 min. In the meantime, the DNA was dephosphorylated in a reaction that included 24 µl genomic DNA (aiming for >210 ng μl^{-1}), 3 µl of 10 × CutSmart(r) buffer, 3 µl nuclease-free H_2O, and 3 µl Quick phosphatase CIP (NEB). Dephosphorylation time was increased to 20 min at 37 °C because it proved to augment the on-target sequences ratio in preliminary tests (results not shown) followed by inactivation at 80 degC for 2 min. The dephosphorylated DNA sample was then subjected to cleavage and dA-tailing in a reaction that comprised 10 µl of Cas9 RNPs, 1 µl 10 mM dATP (NEB) and 1 µl *Taq* DNA polymerase (5,000 U/ml) (NEB). The reaction was incubated at 37 °C for 45 min followed by an inactivation at 72 degC for 5 min. These newly generated A-tailed fragments were ready for ligation of the adapters using the LSK109 or LSK110 sequencing kits (detailed protocol in S1).

Library sequencing

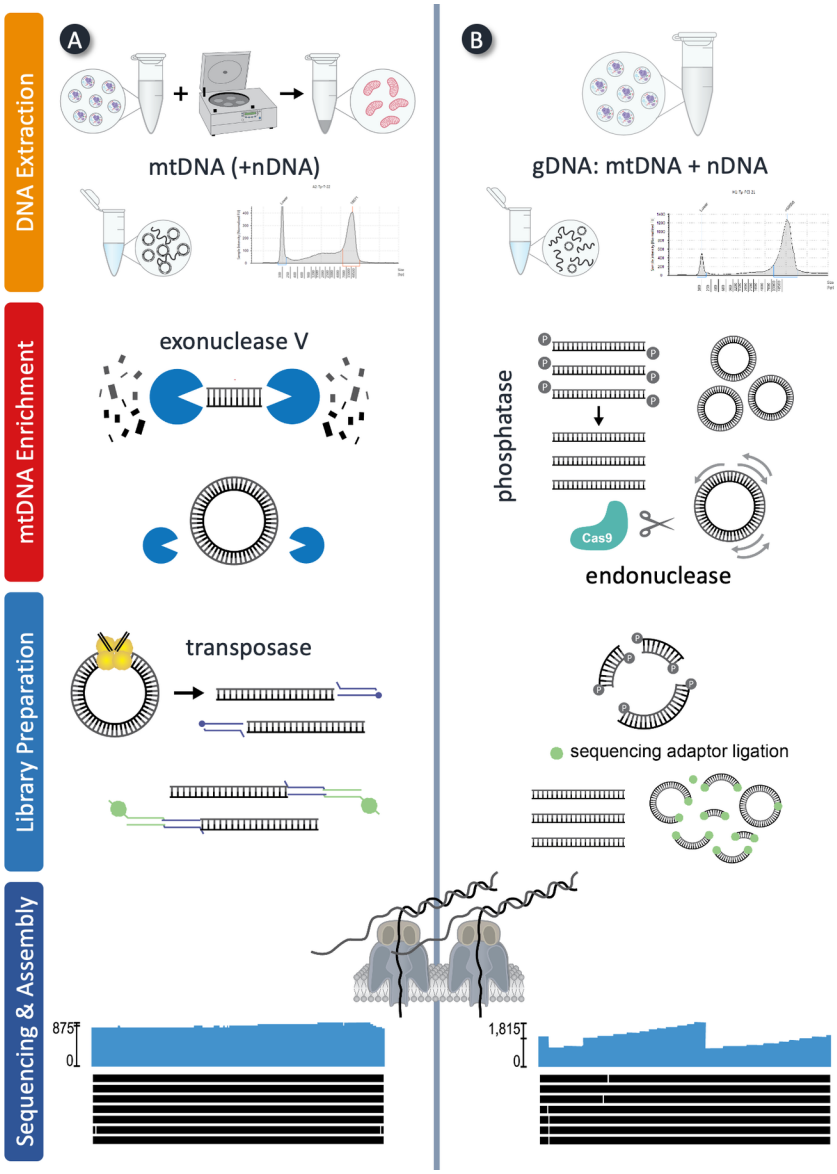
Libraries were run on a MiniION Mk1C (ONT) using a FLO-MIN106 (R9.4.3.), FLO-MIN111 (R10.3) or flongle FLO- FLG001 (R9.4) flow cells following manufacturer's protocol. The sequencing runs were initiated using MinKNOW software (ONT), Raw data were processed with high accuracy basecalling. Sequences with

Qscore>9 were subsequently analyzed with Guppy5 with the 'sup' algorithm, using a custom script in Google Colab. All necessary steps for nDNA depletion, Cas9 targeted sequencing, library preparation, sequencing and analyses were undertaken using a portable setup in a home office during the Covid 19 pandemic.

Mitogenome assembly

Sequences that passed the filter of the Guppy5sup base caller were then assembled de novo using Flye v2.7 (Kolmogorov et al., 2019) using Geneious [flye –nano-raw input_0_Unpaired.fastq –threads 8 –genome-size 0.016m –min-overlap 1000 –iterations 1 –out-dir out]. In the instances in which this de novo assembly failed with the raw data, sequences from 10-18 kbp were extracted and used for de novo assembly in Flye. The resulting circular mitogenome was then used as the reference to map the sequences from its own run to polish the contig and revise the sequencing depth parameters using minimap2 (Li, 2018) within Geneious. Two iterations of the consensus of the contig were built with the highest quality of bases matching at least 60 % and removing the reference genome. This consensus sequence was then annotated using MitoFish (Iwasaki et al., 2013). All consensuses from the different tissues, treatments and runs for chinook salmon were aligned using MAFFT in Geneious and compared. A genome skimming approach was conducted to the two whole genome sequencing runs (C in Table 3) by mapping the sequencing data to the mitogenome of a closely related species using minimap2 (*Osmerus mordax dentex* (MH370836) in the case of*T. pacificus* and *Merluccius merluccius* in the case of*M. productus*) and the resulting consensus sequence without the reference genome was then used as the reference for the second mapping iteration.

Figure 1. Outline of the two enrichment methods from DNA extraction to de novo assembly of the mitogenomes. *A. Mitoenrichment:* Mitochondrial DNA (mtDNA) enrichment by isolating intact mitochondria followed by DNA extraction. Any residual nuclear DNA (nDNA) co-purification is depleted based on the ability of the exonuclease V of cleaving the 5' and 3' termini of any linear DNA (single or double stranded) while leaving circular DNA intact. The circular DNA is then linearized with a transposase that facilitates the addition of the sequencing adaptors. *B. Targeted mitosequencing* : Targeted Sequencing that combines the CRISPR Cas9 nuclease cutting ability directed by a so-called guide RNA or crRNA for targeted scissions of the DNA. The DNA has been previously treated with a phosphatase enzyme that dephosphorylates the free ends of DNA strands preventing any sequencing adapters binding the non-target DNA but allowing the adapters binding to the newly free phosphorylated ends in the selected regions. Both approaches are followed by sequencing on an Oxford Nanopore platform. Illustration by Su Kim, NWFSC/NOAA Fisheries.



Results

DNA quality and quantity

Long fragments of DNA were extracted suggesting very little fragmentation of the DNA with the phenol-chloroform and phase lock method for the fresh *O. tshawytscha* specimen (DIN>9 for gDNA), apart from the fragmentation caused by the transposase activity of the rapid library protocol and the targeted scission of the Cas9 nuclease. Among all the sequences obtained, a great number of complete mitogenome sequences were observed that were simply linearized by the Cas9 or the transposase enzymes. Slightly lower DIN (8.9) was obtained for *T. pacificus* that were kept at -80°C for approximately a month and transported in dry ice for 6 h. Even lower values (<6.5) were obtained for *M. pacificus, T. crenularis, D. theta* and *S. leucopsarus* that were kept at -80°C for a month but transported with ice blocks for 18h. Almost complete degradation of the DNA (DIN<2) was observed on a *T. pacificus* samples that had been always kept at -80°C for more than a year (Table 2, S2).

The mitoenrichment method augmented the ratio of mitochondrial DNA from 2-fold in the case of heart to 26-fold for skeletal muscle (Table 2) and preserved the DNA intact with DIN˜7. Heart and liver produced more mtDNA than muscle, respectively, but also more nDNA and hence the increment of the proportion of target to total DNA is not as pronounced as muscle mitochondrial isolations (Table 2). Liver was not useful in combination with Longmire buffer for the extraction of gDNA because the sample coagulated and made the subsequent extraction steps difficult and an unattainable eluate to work with because the DNA extract was not clean and difficult to pipette. Moreover, the subsequent sequencing run produced very few sequences compared to other runs with no clear size pattern despite having great amounts of DNA and thus liver as a tissue for genomic DNA extraction with Longmire buffer was discarded. However, useful mtDNA was isolated from liver for chinook.

Table 2 DNA extraction comparison by species, tissue type, DNA extraction and integrity (DIN), total DNA quantity, target DNA copy number, target DNA concentration and ratio of target-to-total DNA.gDNA: genomic DNA extraction, mtDNA: mitochondria selection and DNA extraction. The target DNA concentration (only explored for *O. tshawytscha*) is in ng ul^{-1} and target ratio are calculated assuming intact mtDNA of 16,633 bp.

Species	sample ID	tissue	DNA extraction	DIN	total DNA ng ul^{-1}	target DNA co
Oncorhynchus tshawytscha	Ot-PCI-1	muscle	gDNA	9.5	409	8.00E+07
Oncorhynchus tshawytscha	Ot-PCI-H24	heart	gDNA	9	1060	4.14E+08
Oncorhynchus tshawytscha	Ot-C-15	muscle	$gDNA^{+}$	9.4	18.7	8.38E+06
Oncorhynchus tshawytscha	Ot-T-5	muscle	mtDNA	6.9	14.5	7.70E+07
Oncorhynchus tshawytscha	Ot-T-6	muscle	mtDNA	7	10.6	5.40E+07
Oncorhynchus tshawytscha	Ot-T-L7	liver	mtDNA	7.3	310	8.46E+07
Oncorhynchus tshawytscha	Ot-T-H20	heart	mtDNA	9.1	136	1.17E+08
Merluccius productus	Mp-T-L7	liver	mtDNA	3.5	101	
Merluccius productus	Mp-T-H14	heart	mtDNA	3.6	257	
Merluccius productus	Mp-PCI-1	muscle	gDNA	6.3	174.5	
Merluccius productus	Mp-C-3B	liver	$gDNA^{+}$	6.5	500	
Thaleichthys pacificus	Tp-T-22	muscle	mtDNA	6.3	65.9	
Thaleichthys pacificus	Tp-PCI-21	muscle	gDNA	8.9	176	
Thaleichthys pacificus	Tp-PCI-5	muscle	gDNA	1.5^{++}	78.35	
Tarletonbeania crenularis	Tcre-PCI-1	muscle	gDNA	6.2	680	
Diaphus theta	Dthe-PCI-1	muscle	gDNA	6.1	375	
Stenobrachius leucopsarus	Sleu-PCI-1	muscle	gDNA	3.1	690	

$gDNA^{+}$ was extracted using the Nanobind DNA extraction kit (Circulomics) while the rest were extracted using a modified phenol chloroform DNA extraction (PCI) either from crude tissue or from isolated mitochondria (see S3). $^{++}$ Tissue was preserved at -20°C for months and at -80°C for more than a year.

Comparing mitonerichment to targeted mitosequencing

The total number of reads was greater on the mitoenrichment method, however the proportion and number of sequences on target was greater for the targeted mitosequencing method (Table 3, S3, Fig. 2). On the other hand, the read depth was more evenly distributed across the mitogenome with the mitoenrichment method despite the average read depth being greater for the targeted mitosequencing approach. The use of liver increases the ratio of sequences on target for the mitoenrichmnent approach while also increasing the off-target number of sequences, which turns into a greater use of the flow cell. On the contrary, the heart does not perform better than skeletal muscle (Table 3, Fig.3). The more degraded the sample (low DIN) the lower the coverage and the shorter the reads were. Further enrichment was achieved for samples extracted from isolated mitochondria and treated with the Cas9 targeted sequencing method (Fig 3)

Figure 2. Overall representation of all runs and species for the proportion and number of sequences on-target and mean coverage by enrichment method. The targeted mitosequencing method has a greater proportion of sequences on target and deeper coverage overall.

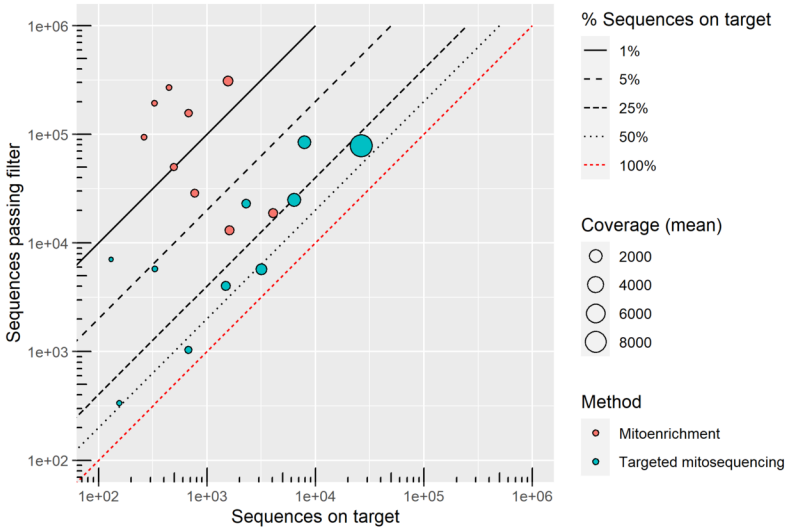
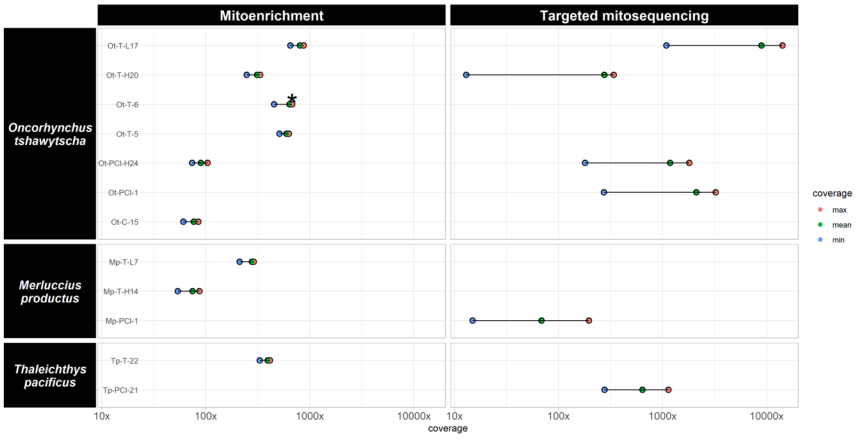


Table 3 Results of the sequencing process and assemblies from three different enrichment methods and different tissue and species. Extraction (method): gDNA - total DNA extraction with phenol-chloroform; mtDNA: extraction with phenol-chlroform from isolated mitochondria. Enrich. method A: mitoenrichment; B: targeted mitosequencing; C: genome skimming after whole genome sequencing. Underlined sample ID indicate those that are available in GenBank with accessions ON005612-ON005619.

Figure 3. Maximum, mean and minimum coverage per treatment (mitoenrichment and targeted mitosequencing), species and sample. Targeted mitosequencing average read depth is overall 10-fold the read depth of the mitoenrichment approach but with more variability within the mitogenome while the mitoenrichment coverage is more uniform. Sample name coding is T: mitochondria selection with ThermoFisher Mitochondria Isolation kit for cultured cells followed by phenol-chloroform DNA extraction; PCI: direct genomic DNA phenol-chloroform extraction; L: liver; H: heart (no notation for skeletal muscle). Note that run Ot-T-6 was performed on a flongle flow cell and run Ot-T-H24 performed on a flongle flow cell that died before two hours of sequencing is not included.



The mitoenrichment method produced fragments of all sizes, including full mitogenomes, because the transposase cleaves the DNA at random places, while the Cas9 in the targeted mitosequencing approach cleaves the DNA at specific sites, hence producing fragments of expected sizes with all the combinations possible with one to five cuts (Fig S3.1). This can be used as an indication of success while monitoring the sequencing run. There are typically peaks of sequences at ˜7.5, 9 or 16kb and the number of sequences passing filter on these peaks can be used as an estimation of the coverage of the final assembly.

The number of sequencing pores is low for these two approaches when compared to Nanopore standards, especially for the targeted mitosequencing method, which can be in the order of 0.3% pores sequencing, while the mitoenrichment can have ca 5% of sequencing pores. This is due to the intense selection of mitochondrial DNA that drastically reduces the number of sequences available for each method, with only very few sequences having sequencing adaptors in the case of the targeted mitosequencing approach. Importantly, these low levels are not an indication of failure, but future improvement of methods should look at optimizing a sequencing run with multiple samples analyzed at once by multiplexing / barcoding samples. When using the flongle flow cell, the targeted mitosequencing run carried took only two hours until functional pores were exhausted; however, the mitoenrichment run lasted for 12 h and produced a mitogenome with 640-fold coverage (Table 2, S2).

Mitogenome assemblies

De novo assemblies of the mitogenome of chinook salmon (*Oncorhynchus tshawytscha*), Pacific hake (*Merluccius productus*), eulachon (*Thaleichthys pacificus*), blue lanterfish (*Tarletonbeania crenularis*), California headlight fish (*Diaphus theta*), and Northern lampfish (*Stenobrachius leucopsarus*) were generated from the Oxford Nanopore data in this project. The *M. productus, T. pacificus* and *T. crenularis*mitogenomes are new contributions to public databases (GenBank Acc. No ON005612-ON005619).

The 11 consensus sequences generated for all the methods optimization tests for chinook salmon are between 16628 and 16634 bp with 99.9% similarity, only differing in homopolymeric regions and follow the canonical vertebrate mitochondrial genome containing 37 genes (13 protein coding, 22 tRNA, and 2 rRNA genes) and 2 noncoding regions (the control region or D-loop and the light strand replication origin or O_L) (Fig 4). These sequences are 10-16 bp shorter than reference genome NC002980 (16,644 bp) (Wilhelm et al., 2003) with 112-134 differences and a pairwise identity ranging from 99.23% (Ot-T-H20-mitosequencing) to 99.33% (Ot-T-L17 and Ot-T-H20 mitoenrichment). The targeted mitosequencing run for the heart sample (Ot-T-H20) was the lowest quality for salmon resulting in a few bp shorter and the most dissimilar sequence to the only reference genome available (NC002980) but also failed to call with confidence a short area near the cutting site of gRNA alias 28 that had shallowed read depth (54-60x) with the strict parameters given. The matching sequences obtained from targeted mitosequencing muscle gDNA and liver mtDNA samples (Ot-PCI-1 and Ot-T-L17), with deepest coverage, was deposited in GenBank as the mitogenome of reference for this individual that only had 3 undetermined nucleotides out of 16633 bp (GenBank Acc. ON005616). All the SNPs found when compared with the reference genome are consistent among the 11 consensus sequences from different tissues, samples, and protocol for the same specimen. Most of the SNPs identified resulted in synonymous substitutions in the coding genes (SNPs on the third position of the codon) and no stops. The exceptions are five non-synonymous substitutions (Ala126Thr, in agreement with (Weitemier et al., 2021), and Ala348Thr on ND2; Ile152Val on COII; Leu116Glu and Met95Thr on ND6). There are also two one-nucleotide insertions and two indels that cause two seven-aminoacid modifications in 519-531aa 571-582aa of ND5. These modifications coincide with other reference sequences for other salmonids (AB252719-AB252722, AY032629- AY032632, LC361126-LC361129) suggesting the reference genome available in GenBank (NC002980) may be incorrect for this region of the ND5 gene. In addition, a previously uncharacterized ND1 aminoacid has been ascertained as Ala118. Despite all the above differences, most of the variation is found in the D-loop with a 98.1 % pairwise identity.

A complete mitogenome was generated for the first time for*Merluccius productus* , for which different haplotypes were found from four different individuals following the canonical vertebrate mitogenome with no gene rearrangements. Three haplotypes obtained from the enrichment methods were deposited in GenBank

(Acc. No ON005613- ON005615). Their length ranges from 16736 to 16775 with the variation found in the non-coding region between genes tRNA-Thr and tRNA-Pro (Fig 4) and at the end of the control region. When the sequences were compared against Atlantic hake (*M. merluccius*) sequence FR751402, the newly identified haplotypes have a shorter non-coding region between genes tRNA-Thr and tRNA-Pro (201 compared to 533 bp) that is missing a 144-bp repetitive region present in the sister species. The genome skimming approach did not produce a complete mitogenome.

A complete mitogenome was generated for the first time for *T. pacificus* for which two haplotypes were identified from two different individuals and three runs, with the two samples from the same individual sequenced using the enriched methods only differing in one nucleotide out of 16,762 bp on a homopolymeric region in the control region (GenBank Acc. No ON005619). The resulting sequence of the genome skimming process was not considered due to its low quality. *T. pacificus* follows a canonical vertebrate mitogenome with no rearrangements or duplications and it contains a repeated motif of a total of 150 bp at the beginning of the control region that was not found on the sequence obtained with the genome skimming approach. Further investigations would be necessary to ascertain the true absence of this repetition cause by either a different haplotype due to a different population origin or bias caused by shallow coverage, short fragment sequencing and degradation and/or the fact that the genome was mapped against a different species.

Three species of myctophids were sequenced, *Tarletonbeania crenularis, Diaphus theta* and *Stenobrachius leucopsarus* , which illustrate the exceptional mitochondrial gene reorganization of this group as described by Poulsen et al. (2013). The three mitogenomes sequenced showed 71.11-75.66% pairwise identity and ca. 4,500-5,500 bp difference in 17.6-18.3kbp. However, no protein coding genes change their order for these three species. The three species present a non-coding indel between tRNA -Leu and ND1 of 39-71 bp and also the typical rearrangement of all myctophids in which the tRNA-Cys (C) and tRNA-Tyr (Y) are switched from the canonical WANCY tRNA-gene order to WANYC (single-letter aminoacid code) and several spacer insertion between genes are found (Fig 4) (Poulsen et al., 2013; Satoh et al., 2016). There is also a longer sequence in the putative origin of replication of the light strand (O_L). *Diaphus theta* and *Stenobrachius leucopsarus* mostly agreed with their publicly accessible references with differences. The sequence generated here for *D. theta* (Acc. No. ON005612) differs in that it is almost 2kbp longer than GenBank Acc. AP012240 (Poulsen et al., 2013) since here we added the complete tRNA-Thr, tRNA-Glu, tRNA-Pro, a novel putative tRNA-Tyr duplication and the D-loop sequences and a 77.1 % pairwise identity at the tRNA-Phe with the rest of the genome only having 46 SNPs. Further analyses are necessary to corroborate the authenticity of the putative duplication of the tRNA-Tyr found in the control region, in agreement with other gene duplications found most often in this area also on birds, reptiles and fishes (Formenti et al., 2021). Nonetheless, it agrees on the Diaphini rearrangement (from IQM to IMQ) in which tRNA-Gln (Q) and tRNA-Met (M) not only switch places but also strands and an extra tRNA-Met pseudogene found between tRNA-Gln and ND2. *S. leucopsarus* complete mitogenome (ON005617) is 2.6 kbp longer than GenBank AP012245 and ND4 gene goes from 98.3% pairwise identity in the first 699 bp to 78.7 in the next 700 bp (158 bp substitutions in 1386 bp resulting in only 13 nonsynonymous aminoacid substitutions). More work is needed to ascertain this remarkable difference in the second half of the ND4 gene. The relocation of tRNA-Glu between tRNA-Thr and tRNA-Pro for *S. leucopsarus* was also observed. *S. leucopsarus* has 44 poly-G in the control region. Lastly, we generated the first publicly available mitochondrial genome for *T. crenularis (* ON005618), which has an insertion of 321 bp between COII and tRNA-Gly.

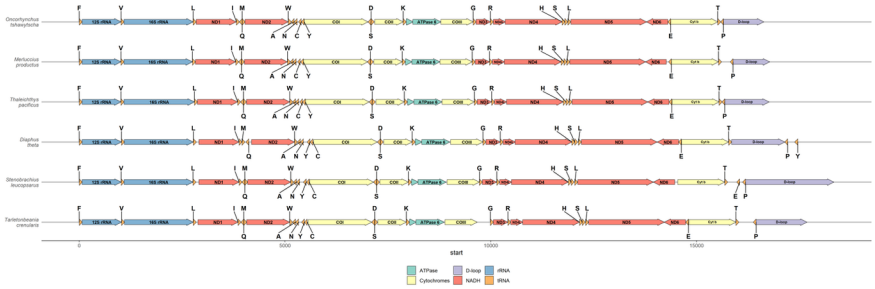
Discussion

This study evaluates simple and cost-effective methods for the discovery of fish mitogenomes for the purpose of building up fish mitochondrial databases. We combined long fragment sequencing and de novo assembly to generate unbiased consensus sequences that do not miss gene duplications, order variations, homopolymeric or tandem repeat regions. We used two mitochondrial DNA enrichment approaches in combination with two different Oxford Nanopore sequencing kits: the mitoenrichment enhances the amount of mtDNA at the extraction and post-extraction steps and can be applied to all fish while the targeted mitosequencing favors the sequencing of mtDNA from previously dephosphorylated genomic DNA with the aid of a CRISPR Cas9

enzyme in combination with five guide RNA sequences that direct the cuts and can be applied to a majority of fish species. Indeed, we obtained novel complete mitogenomes, some following the typical vertebrate mitogenome, some with extremely repetitive segments in the control region and others with gene rearrangements and indels. We generated a new reference mitogenome for chinook (*Oncorhynchus tshawytscha*) and complete mitogenome of another two key species of economical and conservation importance, the Pacific hake (*Merluccius productus*) and eulachon (*Thaleichthys pacificus*). We see the methods tested here as an immediate, speedy, practical sequencing approach that provides certainty and quality to the mitogenome contigs.

Myctophids (lanterfishes and blackchins) are a common finding in our laboratory metabarcoding studies from oceanic samples, being the most abundant family of fishes in the oceans in terms of biomass. Adding resolution further than Myctophidae ASV beyond the phylogenetic effort by Poulsen et al. (2013) is important to resolve species within eDNA metabarcoding work. Myctophiformes is a very diverse family with 250 species described with very divergent larval forms that is far from complete in the molecular databases (Poulsen et al., 2013, also see Fig S3.2). Moreover, very few specimens remain identifiable in survey net hauls because their scales and photophores pattern are often damaged. Getting more voucher specimens sequenced is thus important and de novo assembly of this group of fishes is key because the gene rearrangement made mapping to a close relative mitogenome insufficient, resulting in shorter genomes with missing genes (data not shown). The complete mitogenomes of the blue lanternfish (*Tarletonbeania crenularis*), Northern lampfish (*Stenobrachius leucopsarus*), and California headlighfish (*Diaphus theta*) were generated for the first time in this study. Prior myctophid mitogenomes (Poulsen et al., 2013) were almost complete but were missing the last ca. 2000 bp from the tRNA-Thr, t-RNA-Pro and the control region, due to the difficulty of sequencing such highly polymorphic and repetitive region (e.g. 44 G and >30 repetitions of 18bp on *S. leucopsarus*).

Figure 4. Sequential order of the genes found in the six species studied. The tRNA are noted with their single-letter aminoacid code and there are either they are placed on above or below their line to denote in which strand they are transcribed.



Importantly, the success of the approach for generation of whole mitogenomes is predicated on the availability of high-quality material and proper handling, preserving the integrity and quality of high molecular weight DNA. In this study, we used a modified phenol-chloroform method (Ana Ramón-Laca et al., 2021), directly to crude lysate or after mitochondria isolation, that does not require pipetting steps or manipulation of the DNA to avoid DNA fragmentation. Preventing DNA shearing is crucial for nuclear copies of mitochondrial DNA (NumtS) avoidance and gene order preservation as well as to hamper the linearization of the circular mitochondrial genomes, which would decrease the amount of mtDNA at the nDNA depletion enrichment process. High target DNA (i.e. mitochondrial DNA) concentrations are required to achieve the deep read coverage necessary to overcome the error rates of Oxford Nanopore sequencing. We achieved high DNA concentrations while maintaining good DNA quality with phenol-chloroform extractions with a phase-lock in this study (Table2). However, good DNA handling is not the only parameter that determines the DNA integrity. The level of integrity and the wide range of sample preservation showed in this study demonstrated the speed at which the DNA degrades. We encourage researchers to either preserve samples in dry ice or

at -80°C as soon as possible and to perform the extractions quickly after collection or to preserve the DNA from the first minute in Longmire buffer at room temperature; however, in this case only the targeted mitosequencing approach is possible because the cells and organelles are already lysed and the DNA is in suspension.

The results of the enrichment process (proportion of sequences on target, read depth, fragments length, method to be used) are heavily dependent on the type and level of degradation of the tissue and thus on the quality and quantity of the DNA and enrichment treatment. The genome skimming results demonstrated that the metric of proportion of reads on target does not guarantee a reliable mitogenome, but the coverage and the length of the fragments is what will determine a good result. The amount of mitochondrial DNA found in skeletal muscle (˜0.3% of gDNA) already shows some level of enrichment when compared to the expected 0.1% of the total DNA (Robin & Wong, 1988). The purification of DNA from isolated mitochondria already enriched the proportion of target DNA by two in the case of heart and 26-fold in the case of skeletal muscle. The effect caused by the nDNA depletion in this study, was not measured but empirically observed by a decrease in the amount of total DNA of the sample (results not shown). When comparing the same chinook sample using the two enrichment methods, the targeted mitosequencing approach more than tripled the proportion of reads on target escalating by 67 to 196-fold (Table 2).

To date, the average depth of reads for the mitochondrial genome using PCR-free genome skimming approach with long reads is low [e.g. 9x for the Brazilian buffy-tufted-ear marmoset *Callithrix aurita* , 78x for fresh liver of a rodent of the genus *Melanomys* (Franco-Sierra & Diaz-Nieto, 2020; Malukiewicz et al., 2021)]. Using Nanopore sequencing, the present study achieved only 0.08 % sequences on target and 8.9x coverage (with some areas not covered) for a genome skimming analysis for *Merluccius productus.* Margaryan et al. (2021) obtained an average value of the median depth of coverage for the 192 assemblies of Danish vertebrate species of 1,170.8x, ranging ca. 27– 12,208x, while the fraction of mtDNA reads was around 0.54% ranging from ca. 0.005% to 5.62% depending on the tissue preservation conditions using short-fragment data from genome skimming with HiSeq technology (Illumina). The overall coverage of the contig was dramatically enhanced in this study from hundreds to thousands of reads depending on the tissue, extraction method and treatment (Table 3), with the targeted mitosequencing method providing deepest coverage. The deep coverage results obtained for targeted mitosequencing are, not surprisingly, remarkably higher than other Cas9 targeted assays on nuclear regions of interest. Gilpatrick et al. (2020) obtained a coverage of 80x for a nuclear gene with similar length as the mitogenomes (18kbp), but 680x when targeting the region with multiple cuts. Since the number of available mitochondrial copies is up to 500 times the number of nuclear copies the results are comparable to the latter mentioned study. The cutting sites for the targeted mitosequencing approach are notably less covered than the rest of the mitogenome (File S3.3), but these sites are precisely the least variable regions of the mitogenome among fish. Directional bias is created at the cut site, due to the retention of bound Cas9 ribonucleotide complexes on cleaved DNA fragments that are distal to the PAM site.

While Sanger and mass parallel sequencing of short reads can hinder and obscure challenging tandem repeat regions found in the non-coding regions of the mtDNA of some organisms, such as the tandem repeat insertion in the control region of eulachon or the 44 bp stretch of polyG found on *S. leucopsarus* , in this study (Filipović et al., 2021; Kinkar et al., 2020), long read sequencing can overcome this obstacle. We expect a few scattered errors in the homopolymeric regions in the mitogenomes generated herein (sequencing error rates were [?]2% at the time of this study with Guppy5sup basecalling and R9.4 flow cells), but these are to be fixed as the base calling algorithms evolve or ultimately with the new sequencing kits and flowcells that increase yield and accuracy (SQK-LSK114 and flow cell R10.4) that were just released at the time of finalizing this manuscript, which were shown to reach >99% accuracy. These improvements will presumably allow single nucleotide variants and possible heteroplasmy detection (Keraite et al., 2022). On the other hand, the genomes generated in this study are likely to be more accurate than previous efforts since the methods overcomes the difficulty of sequencing the tandem repeats and long insertions, repetitions and rearrangements that can take place in vertebrate genomes (Formenti et al., 2021), provide read depth, favor long fragments of DNA and augment the proportion of the target. To circumvent sequencing errors, we

recommend read depth of 50-75x for reliable mitogenome generation at the current error rate to get accurate mitogenomes with R9.4 flow cells. Srivathsan et al. (2021) demonstrate that 25-50x for R10.3 flowcells is sufficient and this can only decrease with new improvements.

The mitoenrichment method can be used straight away from mitochondrial isolation material of any fish and is particularly useful in the absence of a reference mitogenome of the species or closely related species of interest to develop guides for the Cas9 for the targeted mitosequencing method. However, the mitochondrial isolation can decrease the mitochondrial DNA amount and there can be nDNA co-purification to some extent. Exonuclease V digestion helps reduce the noise caused by the nDNA, but it is yet incomplete. The rapid sequencing kit (ONT) uses a transposase and a rapid adapter ligation that makes the library preparation brief and straightforward. We speculate that this method will also work on other phyla provided there is enough mitochondrial representation of the target vs microorganisms (i.e. little contamination from symbionts or gut microbiota) as seen in preliminary work performed on bivalves (results not shown, Ramón-Laca personal observation). On the contrary, the Cas9 guides RNAs developed for the targeted mitosequencing method will be useful for most fish and particularly the guides that cut the 12S and 16SrRNA genes. The two guides located on the tRNA-Gly may be less universal since this region is not as conserved and there may be mismatches. Both methods take the same time, around 2.5 h for the library preparation from DNA treatment after the DNA extraction to library load on the flow cell, but the mitoenrichment includes an extra step prior to the DNA extraction to isolate the mitochondria.

The targeted mitosequencing approach tends to shorten the life of the regular flow cells and particularly so in the case of the flongle compared to the mitoenrichment method, presumably due to the presence of the dephosphorylated DNA that may be blocking pores. On the other hand, the mitoenrichment method worked well on a flongle making it an ideal approach for de novo sequencing that will reduce the cost per samples if running them individually almost 10 times compared to the prince of a regular flow cell. However, a refrigerated centrifuge is still needed for the mitochondria isolation and it comes with the disadvantage of the flongle flow cells having a very short shelf-life, which reduces the chances of improvisation, especially in remote places. All the analyses of this study were carried out on a regular computer using standard analysis software within Geneious Primer 2021, with the exception of the Guppy5sup base calling that was performed using Google Colab with a subscription of 10 USD a month, which allowed the use of a high-end GPU that completed basecalling at a rate $>10^6$bp s^{-1}.

At the time of these experiments a barcoding kit or protocol for multiple samples was not available and it was more practical to carry out each mitogenome sequencing in a different run to guarantee good results and reuse the flow cell up to five times (116 USD per mitogenome, sequencing cost without labor and reagents) or use the flongle version of the flow cell for the mitoenrichment workflow (67 USD per mitogenome without labor and reagents). Advancements that make multiplexed runs using barcodes possible with the same yield will reduce the cost significantly while increasing the throughput. (Keraite et al., 2022) run multiple samples by using different guide RNAs for each sample of human origin. However, this is less attainable with de novo sequencing of mitogenomes of fishes given their diversity but could be useful for multiple samples of the same species. Comparatively, the widespread genome skimming approach on a NovaSeq platform (Illumina) can run 200 samples with a cost of sequencing per sample of 140 USD that could be decreased to USD 28 if pooling 1000 samples and aiming for 100x on target read depth (Margaryan et al., 2021). However, this approach is only worth if pooling many different samples, which entails time and a great effort that is likely to involve the coordination of many different institutions and archived samples that may see the sample quality compromised. In addition, intense computational power would be necessary for genome skimming, especially if aiming for deep read approaches to gain coverage that can produce 5 Gb per specimen (Margaryan et al., 2021), which may be impractical for many researchers and laboratories, but also may take weeks to process while the approaches shown here can take two days from DNA extraction to mitogenome annotation. Moreover, typical sequencing lengths of short-fragment platforms from range 150-300 bp and thus this approach is expected to be more prone to include nuclear copies of the mitochondria (Numts) in the contig; resulting in illegitimate or inaccurate sequences. However, we have proven here that in the event of the DNA being degraded (e.g. eulachon Table 3), genome skimming is the only alternative in which case

the researcher will have to assume the potential contamination with Numts that may be overcome by a high proportion of mtDNA that is likely to be due to degradation of linear DNA. This is probably a better approach for archived samples from museums.

Besides supplementing public databases with the standard metabarcoding mitochondrial genes (12S, 16S, COI, and to a lesser extent Cytb or ND2), there are other genes (e.g. D-loop) that may be relevant to closely related species, such as those with more recent evolutionary histories. In addition, the complete mitogenome availability will enhance multiple marker metabarcoding efforts (Leite et al., 2021), providing the resources for population mitogenomics (e.g. hypervariable non-coding region found in this study among three different individuals of Pacific hake) and will provide insightful knowledge about radiation and evolutionary processes with possible gene rearrangement. With deeper sequencing and research, it may even help gain perspective about mitochondrial disorders and diseases of wildlife.

The targeted mitosequencing approach can be further optimized by introducing some of the features used by (Keraite et al., 2022). They pre-enriched by selecting the circular DNA using exonuclease V like we did here for the mitoenrichment approach with a 27-49 % of reads increase (useful for high integrity DNA only) and they also performed a digestion with proteinase K after the cleavage of the DNA to avoid the directional bias seen in our mitosequencing workflow from bound Cas9 enzymes to the DNA that increase the yield of full length reads 2-fold. Further targeted enrichment could possibly be attained at the sequencing step without any further lab-bench effort by using the adaptive sampling capability of the nanopores (Martin et al., 2021; Payne et al., 2021). This process requires reference sequence of the target or related species. We speculate that if all fish mitogenomes are uploaded as references to enrich for, the on-target yield could be improved. However, this tool at the sequencing level requires greater computational power and would need to be validated to ensure it does not bias the results. Alternative methods to enrich the mitochondrial DNA yield could be also pursued using custom myBaits Mito Targeted Sequencing Kit (Arbor biosciences). However, this is a less universal, more complex and longer procedure (Zascavage et al., 2019). Instead, size selection could be achieved with gel extraction (e.g. BluePippin platform), but this instrument was not available for this study and it was not tested.

A great potential advantage of the MinION sequencing platform and associated single flow cell Oxford Nanopore sequencers is the portability outside of a traditional laboratory. The MinION has great portability and the instruments and equipment necessary (e.g. laptop, pipettes, minicentrifuge, small thermal cycler, vortex, plasticware) could also be easily transported in a couple of suitcases. In fact, the experiments in this study were mostly undertaken using a portable setup in a home office during the Covid19 pandemic showing that the lab requirements can be kept to a minimum. Though some of the reagents and the flow cells listed in the methods section would be more limiting as they need to be stored either in the fridge or freezer, these are often readily accessible on research vessels and at remote field stations, or portable refrigeration could be used. In the field, the targeted mitosequencing approach could be modified to be performed in situ, replacing the phenol-chloroform DNA purification with a simpler HMW DNA extraction (e.g. Monarch HMW kit, New England Biolabs), with very little equipment. The mitoenrichment approach uses the Rapid sequencing kit SQK-RAD004 (ONT). The field sequencing kit (SQK-LRK001) is fundamentally the same, but the chemistry is dehydrated to allow its use in the field in the absence of a freezer. This sequencing kit might simplify and allow in situ sequencing; however, further investigations are necessary to determine the yield as it was not tested in this study. The portability of this approach affords the opportunity not only to take these resources to the field, but in doing so, takes advantages of the freshness of samples to yield optimal results, and the production of data in a timely manner.

In summary, the targeted mitosequencing and mitoenrichment approaches, paired with the portable MinION sequencer, gives rapid, cost-effective results for the generation of whole mitogenomes, which are increasingly important for explorations of biological diversity in environmental DNA studies. Mitogenomes can be generated ad-hoc on one to a few samples at a time with little computational effort. Future improvements on samples multiplexing or in sequencing devices will enable high-throughput sequencing of mitogenomes at once. The methods outlined here rely heavily on high quality, (HMW DNA). We strongly recommend the

targeted mitosequencing using the gRNAs designed in this study for all bony fishes, whereas the mitoenrichment would be preferred on distant taxa for which a group-specific nCATS has not been developed. Genome skimming is recommended when only degraded DNA is available (e.g. Museum specimens in ethanol). We believe these approaches will make the generation of reference mitogenomes accessible to many researchers worldwide.

Acknowledgements

Special thanks to Ryan Kelly for facilitating the resources, to Linda Park for comments and conceptualization of the early stages of the project, to Piper Schwenke, Ethan Beyer, Lance Kruzic and Jen Zamon for helping with sample collection. Funding for this project was provided by the National Marine Fisheries Service Genomic Strategic Initiative.

Authors contributions:

* designed research – ARL, KMN, RGS

* performed research – ARL, RGS

* contributed new reagents or analytical tools – ARL, RGS

* analyzed data – ARL, RGS

* wrote the paper – ARL, KMN, RGS

Supplemental material

File S1 – Detailed protocols

File S2 – Data

File S3 – Supplemental methods: attempted experiments and other considerations

Data Accessibility Statement:

Raw sequence reads are deposited in the datadryad.org (A. Ramon-Laca et al., 2022)

Unique haplotype mitogenome data are deposited to NCBI Nucleotide Database (Accessions ON005612-ON005619)

References:

Antunes, A., & Ramos, M. J. (2005). Discovery of a large number of previously unrecognized mitochondrial pseudogenes in fish genomes.*Genomics, 86* (6), 708-717. doi:https://doi.org/10.1016/j.ygeno.2005.08.002

Bourret, V., Albert, V., April, J., Cote, G., & Morissette, O. (2020). Past, present and future contributions of evolutionary biology to wildlife forensics, management and conservation. *Evolutionary Applications, 13* (6), 1420-1434. doi:https://doi.org/10.1111/eva.12977

Brown, K. H. (2008). Fish mitochondrial genomics: sequence, inheritance and functional variation. *Journal of Fish Biology, 72* , 355-374. doi:doi:10.1111/j.1095-8649.2007.01690.x

Dayama, G., Emery, S. B., Kidd, J. M., & Mills, R. E. (2014). The genomic landscape of polymorphic human nuclear mitochondrial insertions.*Nucleic Acids Res, 42* (20), 12640-12649. doi:10.1093/nar/gku1038

Deagle, B. E., Kirkwood, R., & Jarman, S. N. (2009). Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces.*Molecular Ecology, 18* (9), 2022-2038. doi:https://doi.org/10.1111/j.1365-294X.2009.04158.x

Dhorne-Pollet, S., Barrey, E., & Pollet, N. (2020). A new method for long-read sequencing of animal mitochondrial genomes: application to the identification of equine mitochondrial DNA variants. *BMC Genomics, 21* (1), 785. doi:10.1186/s12864-020-07183-9

Dugal, L., Thomas, L., Jensen, M. R., Sigsgaard, E. E., Simpson, T., Jarman, S., . . . Meekan, M. (2021). Individual haplotyping of whale sharks from seawater environmental DNA. *Molecular Ecology Resources, 22* (1), 56-65. doi:https://doi.org/10.1111/1755-0998.13451

Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., . . . Lodge, D. M. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding.*Molecular Ecology Resources, 16* (1), 29-41. doi:https://doi.org/10.1111/1755-0998.12433

Filipović, I., Hereward, J. P., Rašić, G., Devine, G. J., Furlong, M. J., & Etebari, K. (2021). The complete mitochondrial genome sequence of*Oryctes rhinoceros* (Coleoptera: Scarabaeidae) based on long-read nanopore sequencing. *PeerJ, 9* , e10552. doi:10.7717/peerj.10552

Formenti, G., Rhie, A., Balacco, J., Haase, B., Mountcastle, J., Fedrigo, O., . . . The Vertebrate Genomes Project, C. (2021). Complete vertebrate mitogenomes reveal widespread repeats and gene duplications.*Genome Biology, 22* (1), 120. doi:10.1186/s13059-021-02336-9

Franco-Sierra, N. D., & Díaz-Nieto, J. F. (2020). Rapid mitochondrial genome sequencing based on Oxford Nanopore Sequencing and a proxy for vertebrate species identification. *Ecology and Evolution, 10* (7), 3544-3560. doi:https://doi.org/10.1002/ece3.6151

Furlan, E. M., Gleeson, D., Wisniewski, C., Yick, J., & Duncan, R. P. (2019). eDNA surveys to detect species at very low densities: A case study of European carp eradication in Tasmania, Australia. *Journal of Applied Ecology, 56* (11), 2505-2517. doi:10.1111/1365-2664.13485

Gallego, R., Jacobs-Palmer, E., Cribari, K., & Kelly, R. P. (2020). Environmental DNA metabarcoding reveals winners and losers of global change in coastal waters. *Proceedings of the Royal Society B: Biological Sciences, 287* (1940), 20202424. doi:doi:10.1098/rspb.2020.2424

Gan, H. M., Linton, S. M., & Austin, C. M. (2019). Two reads to rule them all: Nanopore long read-guided assembly of the iconic Christmas Island red crab, *Gecarcoidea natalis* (Pocock, 1888), mitochondrial genome and the challenges of AT-rich mitogenomes.*Mar Genomics, 45* , 64-71. doi:10.1016/j.margen.2019.02.002

Gilpatrick, T., Lee, I., Graham, J. E., Raimondeau, E., Bowen, R., Heron, A., . . . Timp, W. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nature Biotechnology, 38* (4), 433-438. doi:10.1038/s41587-020-0407-5

Hartmann, N., Reichwald, K., Wittig, I., Dröse, S., Schmeisser, S., Lück, C., . . . Englert, C. (2011). Mitochondrial DNA copy number and function decrease with age in the short-lived fish Nothobranchius furzeri. *Aging Cell, 10* (5), 824-831. doi:10.1111/j.1474-9726.2011.00723.x

Isokallio, M. A., & Stewart, J. B. (2018). *Isolation of high-quality, highly enriched mitochondrial DNA from mouse tissues.*Retrieved from protocols.io:

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., . . . Nishida, M. (2013). MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular Biology and Evolution, 30* (11), 2531-2540. doi:10.1093/molbev/mst141

Jackman, J. M., Benvenuto, C., Coscia, I., Oliveira Carvalho, C., Ready, J. S., Boubli, J. P., . . . Guimarães Sales, N. (2021). eDNA in a bottleneck: Obstacles to fish metabarcoding studies in megadiverse freshwater systems. *Environmental DNA, 3* (4), 837-849. doi:https://doi.org/10.1002/edn3.191

Jensen, M. R., Sigsgaard, E. E., Liu, S., Manica, A., Bach, S. S., Hansen, M. M., . . . Thomsen, P. F. (2020). Genome-scale target capture of mitochondrial and nuclear environmental DNA from water samples.*Molecular Ecology Resources* (July), 1-13. doi:10.1111/1755-0998.13293

Keraite, I., Becker, P., Canevazzi, D., Frias-López, M. C., Dabad, M., Tonda-Hernandez, R., . . . Gut, I. G. (2022). Novel method for multiplexed full-length single-molecule sequencing of the human mitochondrial genome. *bioRxiv* , 2022.2002.2008.479581. doi:10.1101/2022.02.08.479581

Kinkar, L., Young, N. D., Sohn, W.-M., Stroehlein, A. J., Korhonen, P. K., & Gasser, R. B. (2020). First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach. *PLoS Neglected Tropical Diseases, 14* (8), e0008552. doi:10.1371/journal.pntd.0008552

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

Langlois, V. S., Allison, M. J., Bergman, L. C., To, T. A., & Helbing, C. C. (2020). The need for robust qPCR-based eDNA detection assays in environmental monitoring and species inventories. *Environmental DNA, n/a* (n/a). doi:https://doi.org/10.1002/edn3.164

Leite, B. R., Vieira, P. E., Troncoso, J. S., & Costa, F. O. (2021). Comparing species detection success between molecular markers in DNA metabarcoding of coastal macroinvertebrates. *Metabarcoding and Metagenomics, 5* , e70063.

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

López-Girona, E., Davy, M. W., Albert, N. W., Hilario, E., Smart, M. E. M., Kirk, C., . . . Chagné, D. (2020). CRISPR-Cas9 enrichment and long read sequencing for fine mapping in plants. *Plant Methods, 16* (1), 1-13. doi:10.1186/s13007-020-00661-x

Malukiewicz, J., Cartwright, R. A., Dergam, J. A., Igayara, C. S., Nicola, P. A., Pereira, L. M. C., . . . Roos, C. (2021). Genomic skimming and nanopore sequencing uncover cryptic hybridization in one of world's most threatened primates. *Scientific Reports, 11* (1), 17279. doi:10.1038/s41598-021-96404-6

Margaryan, A., Noer, C. L., Richter, S. R., Restrup, M. E., Bülow-Hansen, J. L., Leerhøi, F., . . . Bohmann, K. (2021). Mitochondrial genomes of Danish vertebrate species generated for the national DNA reference database, DNAmark. *Environmental DNA, 3* (2), 472-480. doi:https://doi.org/10.1002/edn3.138

Martin, S., Heavens, D., Lan, Y., Horsfield, S., Clark, M. D., & Leggett, R. M. (2021). Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *bioRxiv* , 2021.2005.2007.443191. doi:10.1101/2021.05.07.443191

McDonald, T. L., Zhou, W., Castro, C. P., Mumm, C., Switzenberg, J. A., Mills, R. E., & Boyle, A. P. (2021). Cas9 targeted enrichment of mobile elements using nanopore sequencing. *Nature Communications, 12* (1), 3586. doi:10.1038/s41467-021-23918-y

Miya, M. (2022). Environmental DNA Metabarcoding: A Novel Method for Biodiversity Monitoring of Marine Fish Communities. *Annual Review of Marine Science, 14* (1), null. doi:10.1146/annurev-marine-041421-082251

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. doi:doi:10.1098/rsos.150088

Mugnai, F., Meglécz, E., Costantini, F., Abbiati, M., Bavestrello, G., Bertasi, F., . . . Wangensteen, O. S. (2021). Are well-studied marine biodiversity hotspots still blackspots for animal barcoding?*Global Ecology and Conservation, 32* , e01909. doi:https://doi.org/10.1016/j.gecco.2021.e01909

Novak, B. J., Fraser, D., & Maloney, T. H. (2020). Transforming Ocean Conservation: Applying the Genetic Rescue Toolkit. *Genes, 11* (2), 209. doi:10.3390/genes11020209

Ogden, R. (2008). Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. *Fish and Fisheries, 9* (4), 462-472. doi:https://doi.org/10.1111/j.1467-2979.2008.00305.x

Oosting, T., Hilario, E., Wellenreuther, M., & Ritchie, P. A. (2020). DNA degradation in fish: Practical solutions and guidelines to improve DNA preservation for genomic research. *Ecology and Evolution, 10* (16),

8643-8651. doi:https://doi.org/10.1002/ece3.6558

Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B. J., & Loose, M. (2021). Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nature Biotechnology, 39* (4), 442-450. doi:10.1038/s41587-020-00746-x

Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. (2018). Long reads: their purpose and place. *Human Molecular Genetics, 27* (R2), R234-R241. doi:10.1093/hmg/ddy177

Poulsen, J. Y., Byrkjedal, I., Willassen, E., Rees, D., Takeshima, H., Satoh, T. P., . . . Miya, M. (2013). Mitogenomic sequences and evidence from unique gene rearrangements corroborate evolutionary relationships of myctophiformes (Neoteleostei). *BMC Evol Biol, 13* , 111. doi:10.1186/1471-2148-13-111

Ramón-Laca, A., Gallego, R., & Nichols, K. M. (2022). Raw fast5 sequencing data of "Affordable de novo generation of fish mitogenomes using amplification-free enrichment of mitochondrial DNA and deep sequencing of long fragments". *DRYAD* . doi:https://doi.org/10.5061/dryad.jm63xsjdj

Ramón-Laca, A., Wells, A., & Park, L. (2021). A workflow for the relative quantification of multiple fish species from oceanic water samples using environmental DNA (eDNA) to support large-scale fishery surveys. *PLoS ONE, 16* (9), e0257773. doi:10.1371/journal.pone.0257773

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. doi:https://doi.org/10.1111/j.1471-8286.2007.01678.x

Robin, E. D., & Wong, R. (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol, 136* (3), 507-513. doi:10.1002/jcp.1041360316

Rodriguez-Ezpeleta, N., Mendibil, I. a., Alvarez, P., & Cotano, U. (2013). Effect of fish sampling and tissue storage conditions in DNA quality: considerations for genomic studies. *Revista de Investigación Marina, AZTI-Tecnalia, 20* (6), 77-87.

Satoh, T. P., Miya, M., Mabuchi, K., & Nishida, M. (2016). Structure and variation of the mitochondrial genome of fishes. *BMC Genomics, 17* (1), 719. doi:10.1186/s12864-016-3054-y

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* (7), 1004-1013. doi:https://doi.org/10.1002/rra.3610

Schroeter, J. C., Maloy, A. P., Rees, C. B., & Bartron, M. L. (2020). Fish mitochondrial genome sequencing: expanding genetic resources to support species detection and biodiversity monitoring using environmental DNA. *Conservation Genetics Resources, 12* (3), 433-446. doi:10.1007/s12686-019-01111-0

Schultzhaus, Z., Wang, Z., & Stenger, D. (2021). CRISPR-based enrichment strategies for targeted sequencing. *Biotechnol Adv, 46* , 107672. doi:10.1016/j.biotechadv.2020.107672

Sharma, P., & Sampath, H. (2019). Mitochondrial DNA Integrity: Role in Health and Disease. *Cells, 8* (2), 100.

Shaw, J. L. A., Clarke, L. J., Wedderburn, S. D., Barnes, T. C., Weyrich, L. S., & Cooper, A. (2016). Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system.*Biological Conservation, 197* , 131-138. doi:https://doi.org/10.1016/j.biocon.2016.03.010

Shelton, A. O., Kelly, R. P., O'Donnell, J. L., Park, L., Schwenke, P., Greene, C., . . . Beamer, E. M. (2019). Environmental DNA provides quantitative estimates of a threatened salmon species. *Biological Conservation, 237* , 383-391. doi:10.1016/j.biocon.2019.07.003

Shu, L., Ludwig, A., & Peng, Z. (2020). Standards for methods utilizing environmental dna for detection of fish species. In (Vol. 11).

Srivathsan, A., Lee, L., Katoh, K., Hartop, E., Kutty, S. N., Wong, J., . . . Meier, R. (2021). ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. *BMC Biol, 19* (1), 217. doi:10.1186/s12915-021-01141-x

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., . . . Bunce, M. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports, 7* (1), 12240. doi:10.1038/s41598-017-12501-5

Straub, S. C. K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., & Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany, 99* (2), 349-364. doi:https://doi.org/10.3732/ajb.1100335

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. doi:https://doi.org/10.1111/j.1365-294X.2012.05470.x

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. In (Vol. 183, pp. 4-18): Elsevier Ltd.

Wang, S., Yan, Z., Hänfling, B., Zheng, X., Wang, P., Fan, J., & Li, J. (2021). Methodology of fish eDNA and its applications in ecology and environment. *Science of The Total Environment, 755* , 142622. doi:https://doi.org/10.1016/j.scitotenv.2020.142622

Weitemier, K., Penaluna, B. E., Hauck, L. L., Longway, L. J., Garcia, T., & Cronn, R. (2021). Estimating the genetic diversity of Pacific salmon and trout using multigene eDNA metabarcoding. *Molecular Ecology, 30* (20), 4970-4990. doi:https://doi.org/10.1111/mec.15811

Wilhelm, V., Villegas, J., Miquel, Á., Engel, E., Bernales, S., Valenzuela, P. D. T., & Burzio, L. O. (2003). The complete sequence of the mitochondrial genome of the Chinook salmon, *Oncorhynchus tshawytscha* . *Biological Research, 36* , 223-231.

Zascavage, R. R., Thorson, K., & Planz, J. V. (2019). Nanopore sequencing: An enrichment-free alternative to mitochondrial DNA sequencing. *Electrophoresis, 40* (2), 272-280. doi:10.1002/elps.201800083