Conservation Genet Resour (2018) 10:483–486
DOI 10.1007/s12686-017-0854-6

TECHNICAL NOTE

Demonstration of a basic mitochondrial enrichment method to produce the complete mitochondrial genome sequence of the endangered North Atlantic right whale (*Eubalaena glacialis*)

J. S. Allwood[1] · M. K. Scheible[1] · S. A. Faith[1]

Received: 13 July 2017 / Accepted: 28 August 2017 / Published online: 4 September 2017

Abstract The North Atlantic right whale is one of only three species within the *Eubalaena* genus, a group heavily targeted by the whaling industry prior to the twentieth century. All three species (*Eubalaena australis, Eubalaena japonica* and *Eubalaena glacialis*) are CITES listed (Appendix I), with the latter two species listed as endangered (IUCN, v 2017-1). The mitochondrial genome of the first two species have been sequenced and are publicly available, while *E. glacialis* was unavailable. Here we present the complete mitochondrial genome sequence of *E. glacialis*, while also detailing a straight forward mitochondrial enrichment method that facilitates analysis using next generation sequencing. This method simultaneously allows for efficient sequencing of the mitochondrial genome, while demonstrating quality sequence coverage that is even across the genome. This enrichment illustrates a marked improvement in sequence depth compared to that achieved when performing whole genome sequencing using a standard DNA extraction.

Keywords Mitochondrial enrichment · *Eubalaena glacialis* · Mitochondrial genome · Next generation sequencing · Bioinformatics

Blood preserved in EDTA from a North Atlantic right whale (New England Aquarium Identification: Eg#3710) was provided by the Center for Marine Sciences and Technology (CMAST, NCSU, NC, USA; Case report CALO2009-01, Harms et al. 2014). The standard DNA extraction [extracted previously using QIAamp Mini Blood Kit (Qiagen, Hilden, Germany)] used the same stored starting material as the enriched extraction undertaken here. The mitochondrial enrichment method was adapted from Ahmad et al. (2007) and Coutinho et al. (2017) for smaller starting volumes of blood (milliliters to microliters), to allow for greater application in wildlife research studies. TKM1 buffer was added in a 1:1 ratio to blood (250 µl each), with 1% volume of IGEPAL, and incubated at room temperature for 10 min, followed by a 20 min centrifugation at 800×*g*. Supernatant was retained and this step was repeated using the same quantities and centrifugation conditions applied as above. The retained supernatant from both TKM1 steps were combined and centrifuged for 20 min at 15,000×*g*. The resulting pellet was briefly washed twice in TKM1 and suspended in 500 µl TKM2 with 100 µl of 10% SDS and incubated overnight at 55 °C. Proteins were salted out with the addition of 400 µl of 3M NaCl, followed by centrifugation for 20 min at 12,000×*g*. The supernatant was halved and transferred to duplicate microcentrifuge tubes. Absolute ethanol was added to each tube at twice the supernatant volume. Samples were then centrifuged for 15 min at 30,000×*g*. Supernatant was removed and the remaining pellet was washed twice in 70% ethanol and hydrated in 75 µl 1× TE buffer (per tube).

✉ J. S. Allwood
jallwoo@ncsu.edu

M. K. Scheible
mkscheib@ncsu.edu

S. A. Faith
safaith@ncsu.edu

[1] Department of Molecular Biomedical Sciences and The Forensic Sciences Institute, North Carolina State University, 1060 William Moore Dr., Raleigh, NC 27607, USA

Table 1 Library and assembly details of both extraction methods

	Standard		Enriched	
Cycles	5	10	5	10
Quant (nM)[a]	2.60	106.07	1.91	138.84
Ave. size	566	563	593	548
Reads	2,417,850	1,349,256	1,921,581	1,318,051
Mapped	2214	1152	11,040	7764
Ave. coverage	40	21	200	141

[a]Library concentration prior to normalisation. All libraries were normalised to 2 nM for sequencing

Both standard and mitochondrial enriched DNA extractions underwent size selection prior to sequence library preparation by first shearing the DNA using the S220 Focused-Ultrasonicator (Covaris, MA, USA). Size selection targeting 400 base pairs (bp) was then performed in duplicate (A and B samples per extraction type) using the Pippin Prep (Sage Science, MA, USA) and quantified using the 2200 TapeStation (Agilent Technologies, CA, USA). Sequence libraries were prepared using the KAPA Hyper Prep Kit (KAPA Biosystems, MA, USA) following the manufacture's guidelines. To achieve appropriate input quantity, each library was PCR enriched, with one replicate of each extraction type undergoing either five or ten cycles of amplification. All four libraries were quantified using the KAPA Library Quantification Kit (KAPA Biosystems) and normalised to an equimolar pool for loading onto a 300 cycle mid-output kit and sequencing on the MiniSeq System (Illumina, CA, USA).

Output files were assessed for quality using FastQC Version 0.11.5 (Andrews 2016) with each sample producing average read lengths as expected (majority ~300 bp) and of high quality (average Phred score >35; no poor quality flags). Sequences were assembled using CLC Genomics Workbench 9.5 software (Qiagen) and mapped against the reference mitochondrial genomes of the two

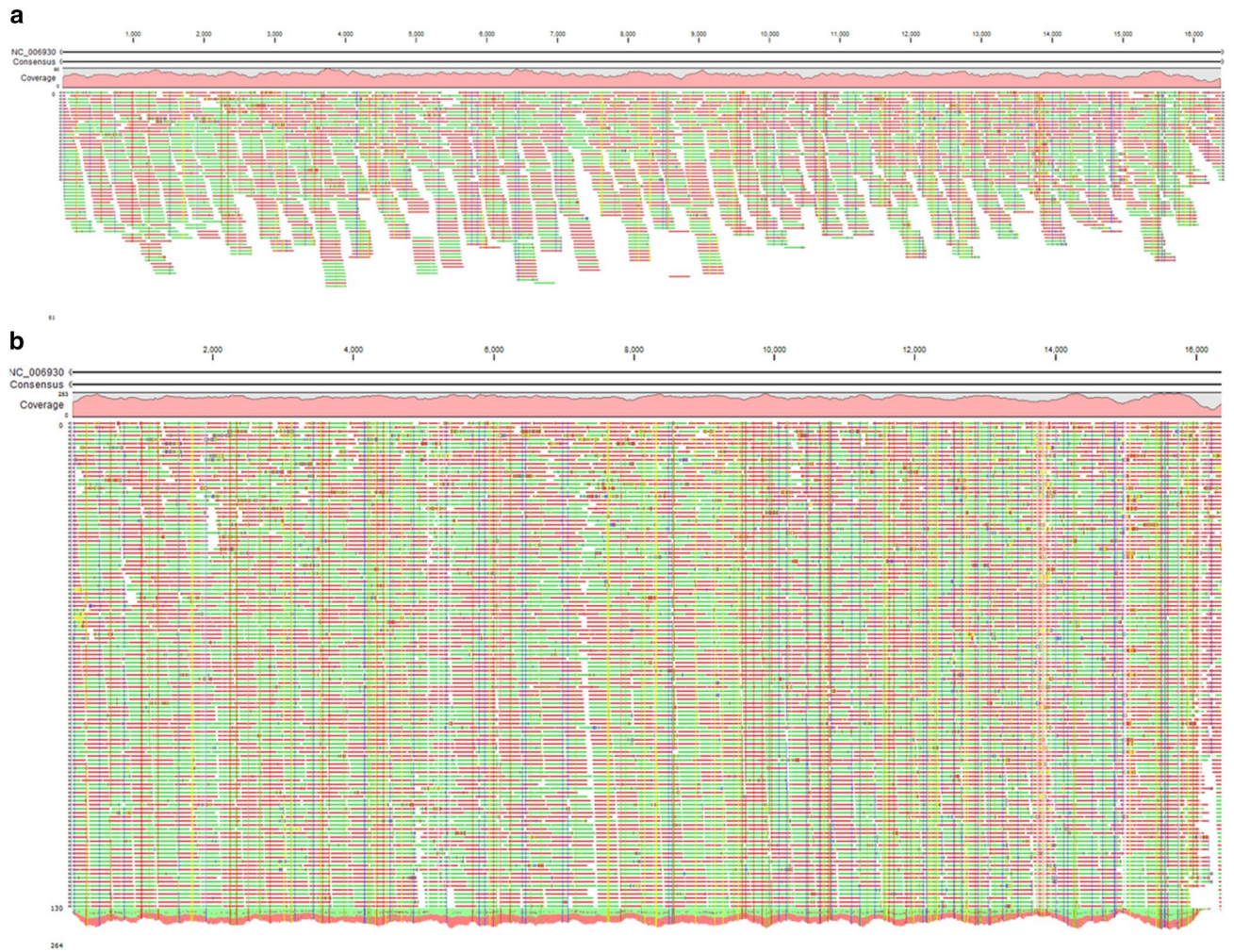

Fig. 1 Mapping comparison of both standard (**a**) and enriched (**b**) five cycle amplification libraries

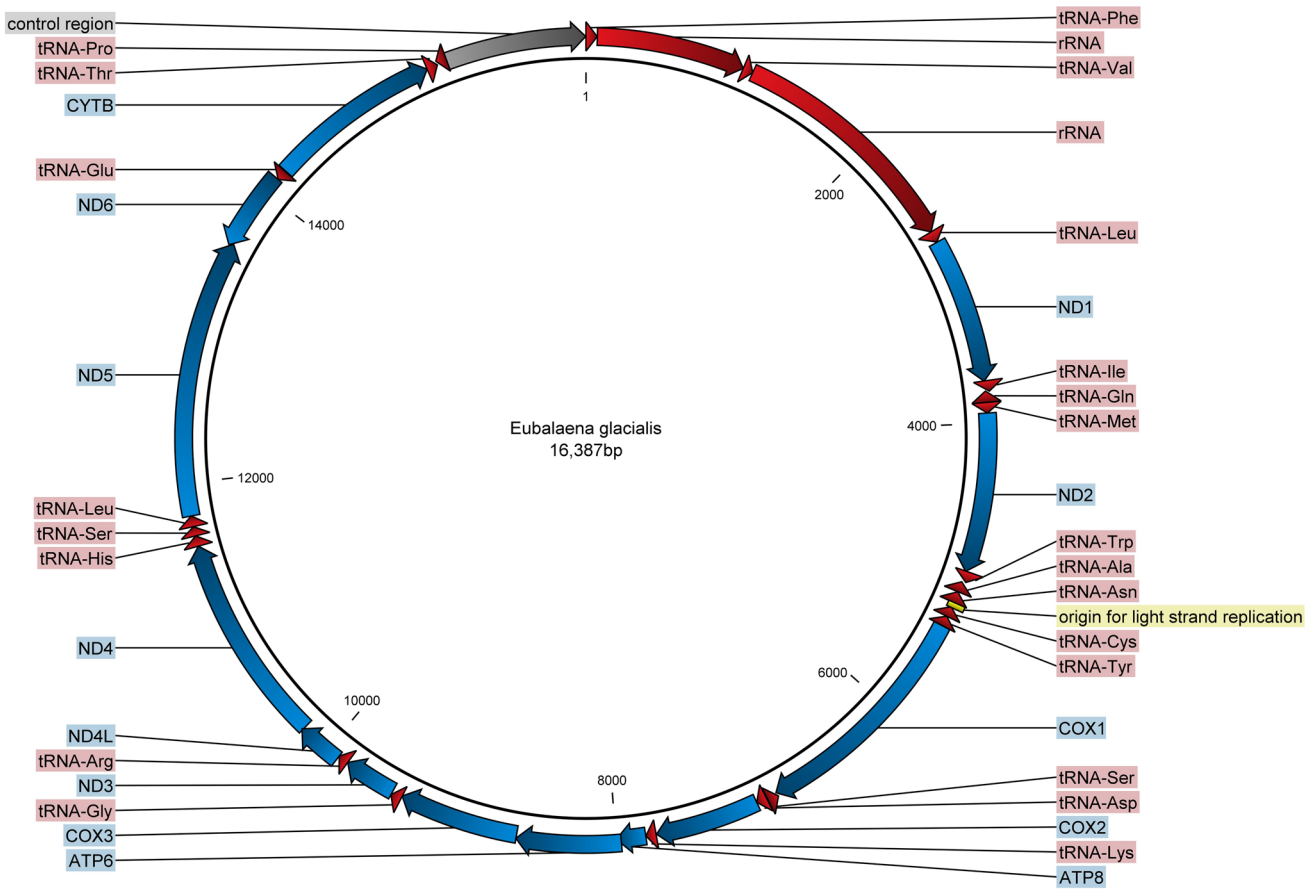

Fig. 2 Map of the mitochondrial genome for *E. glacialis*. Gene annotations as mapped against *E. australis* (GenBank: NC_ 006390) using CLC Genomics Workbench 9.5 Software

congener species *E. australis* and *E. japonica* (Genbank NC_006390-006391, respectively). Consensus sequence for the *E. glacialis* mitochondrial genome did not vary in composition, regardless of which species was utilised as the reference for sequence mapping. Assembly and coverage information for both standard and enriched DNA extractions are detailed in Table 1. Coverage depth of the enriched sample compared to the standard DNA extraction was vastly improved, with a fivefold increase in depth observed at five cycles of library amplification (Fig. 1). Coverage uniformity across the genome decreased with the increase in library cycling to ten cycles (data not shown) in both extraction types, with coverage becoming uneven, perhaps indicating amplification stochasticity.

The enriched five cycle amplification library was selected to ascertain the final mitochondrial genome sequence for *E. glacialis*. This assembly illustrated clear consensus of mapped sequences across the genome, with any sequence variants against the reference genome supported by substantial coverage depth.

The resulting mitochondrial genome for *E. glacialis* (GenBank: MF459656; Fig. 2) was 16,387 bp in size and was annotated using the reference sequence annotations (GenBank: NC_006390). This sequence was aligned against both other *Eubalaena* species as well as with *Balaena mysticetus*, representative of the most closely related cetacean species to the *Eubalaena* genus. Sequence comparison of *E. glacialis* to *E. japonica* identified 232 differences, while comparison to *E. australis* revealed 201 differences, indicating 98.6 and 98.8% similarity to *E. glacialis* respectively. In comparison, *B. mysticetus* showed 95.4, 95.3 and 95.5% percent identity to *E. glacialis, E. japonica* and *E. australis* respectively.

The ability to target mitochondria will increase the efficiency of sequencing projects by allowing for greater pooling capacity without compromising depth of coverage, thereby facilitating the sequencing of multiple samples in a run. This enrichment method becomes more suitable for wildlife applications, and what is typically precious material, through the adaption of using smaller starting volumes. Here we present the final product of this method in the form of the complete mitochondrial genome of the endangered North Atlantic right whale. This genome contributes to the growing need for available reference data for conservation purposes.

Acknowledgements Our thanks to Dr. Craig Harms and the Center for Marine Sciences and Technology (CMAST, North Carolina State University, NC, USA) for kindly providing the sample material obtained from the CMAST marine mammals tissue databank and to all those that assisted with original sample collection. Samples were collected during stranding response authorised by the National Marine Fisheries Service Marine Mammal Health and Stranding Response Program, Southeast Region Stranding Network, Marine Mammal Stranding Agreement with the University of North Carolina Wilmington and Designee North Carolina State University.

Funding Stranding response and sample collection was assisted with funding from the John. H. Prescott Marine Mammal Stranding Assistance Grant Program (NA09NMF4390226, NA08NMF4390590). Sequencing work was funded through the North Carolina State University Chancellor's Faculty Excellence Program.

References

Ahmad S, Ghosh A, Nair DL, Seshadri M (2007) Simultaneous extraction of nuclear and mitochondrial DNA from human blood. Genes Genet Syst 82(5):429–432

Andrews S (2016). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Coutinho E, Batista C, Sousa F, Queiroz J, Costa D (2017) Mitochondrial gene therapy: advances in mitochondrial gene cloning, plasmid production, and nanosystems targeted to mitochondria. Mol Pharm 14(3):626–638

Harms CA, McLellan WA, Moore MJ, Barco SG, Clarke EO, Thayer VG, Rowles TK (2014) Low-residue euthanasia of stranded mysticetes. J Wildl Dis 50(1):63–73

The IUCN Red List of Threatened Species. Version 2017-1. http://www.iucnredlist.org. Accessed 7 July 2017