Species identification and mitochondrial genomes of ancient fish bones from the Riverine Kachemak tradition of the Kenai Peninsula, Alaska

Alida de Flamingh\textsuperscript{a}, Elizabeth K. Mallott\textsuperscript{b}, Alfred L. Roca\textsuperscript{a,c,d}, Alan S. Bora\textsuperscript{e} and Ripan S. Malhi\textsuperscript{a,c,f}

\textsuperscript{a}Program in Ecology, Evolution and Conservation Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA;\textsuperscript{b}Department of Anthropology, University of Alaska, Anchorage, AK, USA;\textsuperscript{c}Department of Anthropology, Northwestern University, Evanston, IL, USA;\textsuperscript{d}Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA;\textsuperscript{e}Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA;\textsuperscript{f}Department of Anthropology, University of Alaska, Anchorage, AK, USA;\textsuperscript{g}Department of Anthropology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

\textbf{ABSTRACT}

Seven fish vertebrae were chosen for analysis from the 49-KEN-147 archaeological site in the Kenai Peninsula, Alaska. Mitochondrial DNA analysis of the ancient fish bones revealed that they were from sockeye and coho salmon. Here, we report the ancient mitochondrial genomes for three sockeye salmon and one coho salmon fish bone.

Seven fish vertebrae were chosen for analysis from the 49-KEN-147 archaeological site (60°28′53.98″N, 151°7′17.96″W; Reger 1998) in the Kenai Peninsula, Alaska. The bones are from a within-house conical storage pit (50 cm × 30 cm) of the Riverine Kachemak tradition that contained 1166 ribs, fins, and vertebrae. Wood structural material from the house dated between 1890 ± 60 and 1570 ± 30 radiocarbon years while the fish bones dated considerably older at 2520 ± 30 radiocarbon years. The discrepancy between the terrestrial dates of wood and the anadromous salmon dates could be due to marine reservoir effect although the almost 1000-year difference is much larger than previously reported for the North Pacific (Edinborough et al. 2016). Ancient DNA analysis provides an effective way of determining the species identity of each of these fish bones (Grier et al. 2013). Here, we used molecular techniques to extract DNA and sequence whole mitochondrial genomes for these samples.

DNA extractions were carried out in a dedicated clean room at the ancient DNA laboratory in the Carl R. Woese Institute for Genomic Biology of the University of Illinois at Urbana-Champaign (UIUC) following a protocol optimized for ancient DNA extraction (Cui et al. 2013). Unused fishbone fragments and extracted DNA samples have been placed in permanent storage in the ancient DNA laboratory (Carl R. Woese Institute for Genomic Biology, Urbana, IL). DNA extracts were tested for endogenous fish DNA by amplifying two short DNA sequences from the cytochrome b (CYTB) and D-loop regions on the mitochondrial genome. These primers have previously been used to identify different fish species from bones at archaeological sites (Yang et al. 2004). We were able to amplify the CYTB region for six and the D-loop region for all seven fish bones (GenBank accession numbers: CYTB MG993156–MG993161, D-loop MG993149–MG993155). We compared these sequences to previously reported sequences on GenBank using basic local alignment search tool (Altschul et al. 1990). The CYTB region for five of the six sequences and the D-loop region for six of the seven sequences matched sockeye salmon (\textit{Oncorhynchus nerka}). A single fish bone had CYTB and D-loop sequences that matched coho salmon (\textit{Oncorhynchus kisutch}). All but one of the corresponding GenBank sequences were exact matches (100% identity) for the entire fish bone sequence (100% coverage). One D-loop sequence that matched sockeye salmon had a 99% identity for 100% coverage due to a single indel.

To reconstruct the mitochondrial genome, we selected four samples that had distinct haplotypes for the short mtDNA regions: three of the sockeye salmon and the single coho salmon. We used a NEBNext Ultra DNA Library Prep Kit for Illumina in combination with NEBNext single index primers to build genomic libraries for each of these fish bones. Libraries were pooled and submitted for HiSeq 4000 sequencing at the UIUC Core Sequencing Facility, which produced over 215 million reads for the pool.

Reads were de-multiplexed and sequences were trimmed using AdapterRemoval (Lindgreen 2012) with the minimum sequence length set to 25 bp. Trimmed sequences were aligned to the \textit{Oncorhynchus kisutch} and \textit{Oncorhynchus nerka} mitogenomes (NC009263 and NC008615, respectively) using
bowtie2 (Langmead and Salzberg 2012) with the local alignment option and capping fragment length at 1000 bp. Aligned sequences were transformed to BAM format in SAMtools v. 1.1 (Li et al. 2009), removing unmapped reads and reads with a quality score less than 30. Aligned files were sorted and indexed using SAMtools v. 1.1 and PCR duplicates were removed in SAMtools v. 0.1.19 (Li et al. 2009). Consensus sequences were generated from the de-duplicated alignment files in Geneious R7 (www.geneious.com). Variants were called in SNVer (Wei et al. 2011) with a mapping quality of 30, a base quality of 20, an alternate/reference ratio of 0.9, and an alternate threshold count of 200. Ancient DNA damage patterns were verified by aligning trimmed reads to the O. kisutch or O. nerka mitogenomes with BWA (Li and Durbin 2010) and quantifying damage in mapDamage2 (Jónsson et al. 2013).

A maximum likelihood (ML) phylogeny was inferred using RaXML v. 8 (Stamatakis 2014); it included mitogenomes from the four fish bones and 10 previously published reference genomes (accession numbers provided in Figure 1) for all North American salmon species, which included Pacific salmon (chinook, chum, coho, pink, and sockeye) and Atlantic salmon. Sequences were aligned using Geneious R7 and the ML tree was drawn using FigTree V 1.4.2 (Rambaut and Drummond 2009). ML analysis parameters included 1000 bootstrap pseudoreplicates and the GTR + Gamma evolutionary model. The ML analysis supported the results of the short mtDNA regions, as three of the mitogenomes grouped with previously reported sockeye salmon and the fourth mitogenome with a previously reported coho salmon mitogenome. The species identifications are consistent with interpretations of a salmon-intensive subsistence strategy of the Riverine Kachemak tradition targeting sockeye salmon and including other species such as coho salmon (Reger 1998).

**Disclosure statement**

The authors report no conflicts of interest.

**Funding**

AdeF was supported by the Cooperative State Research, Education, and Extension Service, US Department of Agriculture, under project number [ILLU 875-952 and ILLU-538-939]. ASB was supported by Kenai-Kachemak Fund, Director's Research Funds, and Kenai Peninsula College.

**References**

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Cui Y, Lindo J, Hughes CE, Johnson JW, Hernandez AG, Kemp BM, Ma J, Cunningham R, Petzelt B, Mitchell J. 2013. Ancient DNA analysis of mid-Holocene individuals from the Northwest Coast of North America reveals different evolutionary paths for mitogenomes. PLoS One. 8:e66948.

Edinborough K, Martindale A, Cook GT, Supernant K, Ames KM. 2016. A marine reservoir effect: Δ R value for kitandach, in Prince Rupert Harbour, British Columbia, Canada. Radiocarbon. 58:885–891.

![Figure 1. A maximum likelihood phylogeny of four ancient fish mitochondrial genomes and 10 previously reported mitochondrial reference genomes from all North American salmon species (labels correspond to GenBank accession numbers) showed that three of the ancient fish bones group with sockeye salmon and a fourth fishbone groups with coho salmon. Bootstrap support is indicated to the right of each node; tree is midpoint rooted; scale bar represents genetic change as the mean number of nucleotide substitutions per site. GenBank accession numbers for the ancient fish bones: FB1 MG993162, FB2 MH003639, FB4 MH003640, and FB7 MH003641.](image-url)
Grier C, Flanigan K, Winters M, Jordan LG, Lukowski S, Kemp BM. 2013. Using ancient DNA identification and osteometric measures of archaeological Pacific salmon vertebrae for reconstructing salmon fisheries and site seasonality at Dionisio Point, British Columbia. J Archaeol Sci. 40:544–555.

Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. 2013. mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 29:1682–1684.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 26:589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078–2079.

Lindgreen S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Res Notes. 5:337.

Rambaut A, Drummond A. 2009. FigTree v1. 3.1. Institute of Evolutionary Biology, University of Edinburgh.

Reger DR. 1998. Archaeology of the northern Kenai Peninsula and upper Cook Inlet. Arctic Anthropol. 35:160–171.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.

Wei Z, Wang W, Hu P, Lyon GJ, Hakonarson H. 2011. SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. Nucleic Acids Res. 39:e132–e132.

Yang DY, Cannon A, Saunders SR. 2004. DNA species identification of archaeological salmon bone from the Pacific Northwest Coast of North America. J Archaeol Sci. 31:619–631.