



HAL
open science

Decoding ancient genomes: Genomics approaches and innovative species recognition pipeline for diadromous fish

Patrick Jacques, Sarah Maman, Régis Debruyne, Brice Ephrem, Aurélie Manicki, Christophe Klopp, Myriam Sternberg, Gaston Rognon, Léonard Ransan, Chloé Bellanger, et al.

► To cite this version:

Patrick Jacques, Sarah Maman, Régis Debruyne, Brice Ephrem, Aurélie Manicki, et al.. Decoding ancient genomes: Genomics approaches and innovative species recognition pipeline for diadromous fish. 20. *Portugaliæ Genetica: DNA - Ancient and New*, Mar 2024, Porto, Portugal. 2024. hal-04527107v3

HAL Id: hal-04527107

<https://hal.inrae.fr/hal-04527107v3>

Submitted on 29 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Decoding ancient genomes: Genomics approaches and innovative species recognition pipeline for diadromous fish

Patrick Jacques^{1*}, Sarah Maman^{2*}, Régis Debruynne³, Brice Ephrem⁴, Aurélie Manicki¹, Christophe Klopp⁵, Myriam Sternberg⁶, Gaston Rognon⁷, Léonard Ransan⁷, Chloé Bellanger⁷, Joelle Chat¹, Philippe Béarez³, Natacha Nikolic^{1,8}

¹INRAE – Institut National de Recherche pour l’Agriculture, l’alimentation, et l’Ecologie, ECOBIOP

²INRAE, SIGENAE, GenPhySE, Toulouse, France

³CNRS/MNHN, UMR 7209, Paris, France

⁴CNRS, UMR 6566, CReAAH, Rennes, France

⁵INRAE, Sigenae, MIAT

⁶CNRS, UMR 7299, Aix-en-Provence, France

⁷Université Paul Sabatier, France

⁸CRBE, Aquaeco, Université de Toulouse, France

(*) Co-first authors

patrick.jacques@inrae.fr

Partners: Barbaza Michel (a), Clavel Benoît (b), Dachary Morgane (a,c), Daujeard Camille (d), Gey Delphine (e), Guillaud Emilie (b), Hamilton-Dyer Sheila (f), Harland Jennifer (g), Hinguant Stéphan (h), Horad-Herbin Marie-Pierre (i), Hummel Esmee (j), Van Der Jagt Inge (j), Jonsson Leif (k), Madelaine Stéphane (l), Makowiecki Daniel (m), Perrin Thomas (a), Primault Jérôme (a,n), Quinlan Liz (o), Schmölcke Ulrich (p)

(a) UMR 5608 TRACES, Toulouse, France

(b) UMR 7209 AASPE, Paris, France

(c) SRA Nouvelle Aquitaine, France

(d) UMR 7194 HNHP, Paris, France

(e) UMS 2700, MNHN, Paris, France

(f) Bournemouth University, U.K.

(g) Archaeology Institute, University of the Highlands and Islands, Orkney, Scotland

(h) INRAP Grand Ouest, France

(i) UMR 7324 CITERES, Tours, France

(j) Cultural Heritage Agency of the Netherlands, Netherlands

(k) University of Gothenburg, Gothenburg Natural History Museum, Sweden

(l) Musée National de la Préhistoire, Les Eyzies, France

(m) Institute of Archaeology, Poland

(n) UMR 7041 ArScAn, Nanterre, France

(o) University of York, U.K.

(p) The Munich State Zoological Collection, Germany

ABSTRACT

Understanding the evolution of wild species is crucial for developing effective population management strategies and predicting their future trajectories. In my doctoral research, we opted for genomics approaches to explore the demographic history of two iconic diadromous fishes: *Anguilla anguilla* (European eel) and *Salmo salar* (Atlantic salmon). We aim to elucidate their adaptive responses to environmental changes. Setting this project apart is the sequencing and genotyping of both archaeological (aDNA) and modern DNAs samples. Our study encompasses archaeological samples from various locations across Europe, spanning a temporal range of 17,000 years – from the Paleolithic to the modern era. To accurately assign species to each sample, we developed DeFiS (*Detect Fish Species*), a pipeline relying on reference genomes and samples to analyze k-mers: fixed-size nucleotide sub-strings. The list of k-mers and their occurrences from specific dictionaries for both reference genomes and samples. By comparing the dictionaries of the samples with those of the references, we could assign each sample to a specific species. We are currently validating the pipeline using freely available modern samples of all European diadromous fish species found in public databases (NCBI, Essembl, and scientific paper). The aim is to provide a valuable resource for biological scientists. Our method offers a significant advancement in the species identification process for aDNA research. By accelerating and refining this crucial step we aim to contribute to future aDNA-based investigations by facilitating the work of biologists studying wild species evolutions.

Decoding ancient genomes: Genomics approaches and innovative species recognition pipeline for diadromous fish

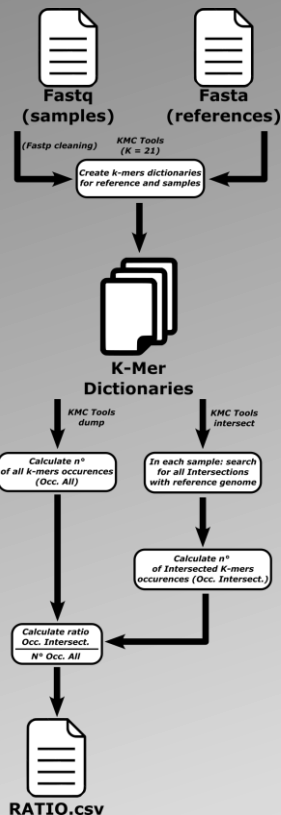
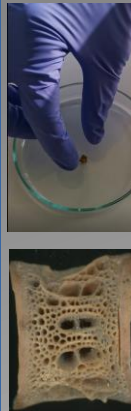
Patrick Jacques^{1*}, Sarah Maman^{2*}, Régis Debruyne³, Brice Ephrem⁴, Aurélie Manicki¹, Christophe Klopp⁵, Myriam Sternberg⁶, Gaston Rognon⁷, Léonard Ransan⁷, Chloé Bellanger⁷, Joelle Chat¹, Philippe Béarez³, Natacha Nikolic^{1,8}

¹INRAE, ECOBIOP ; ²INRAE, SIGENAE, GenPhySE ; ³CNRS/MNH, UMR 7209 ; ⁴CNRS, UMR 6566, CRéAAH ; ⁵INRAE, Sigenae, MIAT ; ⁶CNRS, UMR 7299 ; ⁷Université Paul Sabatier ; ⁸CRBE, Aquaeco. (* Co-first authors)

Introduction and study context

Understanding the evolution of wild species is crucial for developing effective population management strategies and predicting their future trajectories. Ancient DNA (aDNA), DNA collected from archaeological samples, has the potential to serve as a window into the demographic history of species. But **Ancient DNA suffers from three major issues** : it is fragmented, degraded, and contaminated.

Taxonomic assignation of aDNA samples is thus a complex affair. To meet this challenge, we are developing **DeFiS**, a pipeline for aDNA taxonomic assignation based on **K-mers** : fixed-size nucleotide sub-strings.



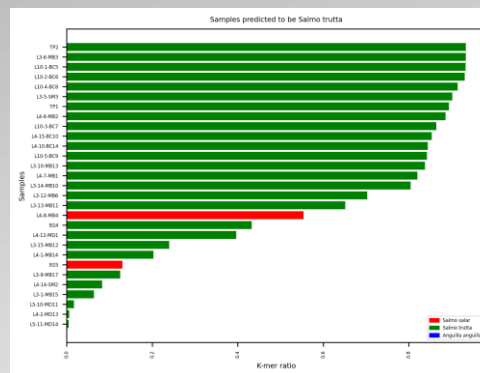
Method

A dictionary is a list of **k-mers occurrences** found in a genome. We can link a reference genome to our sample thanks to the **DeFiS pipeline** which can determine samples' species with a **k-mers ratio**:

$$\frac{Nb_{Occ^*} \text{ Kmers Intersect}^*}{Nb_{Occ^*} \text{ Kmers total}}$$

The highest ratio shows the sample' species.

At the end, the tool will enable the **determination of the taxonomy** of an ancient or modern DNA sequence by comparing it with the dictionary of reference databases



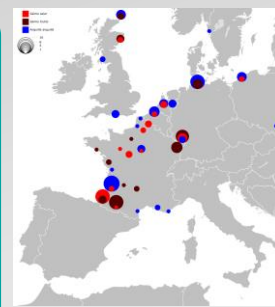
Results

Our method offers a significant advancement in the species identification process for aDNA research on **diadromous fishes**. By accelerating and refining this crucial step we aim to contribute to future aDNA-based investigations by facilitating the work of biologists studying wild species evolution.

We are currently validating the pipeline using freely available modern samples of all European diadromous fish species found in public databases (NCBI, Ensembl, and scientific papers). The aim is to provide a **valuable resource for biological scientists**.

On the left: one example of results regarding the identification of presumed diadromous fish species by archaeoichthyologists, such as Atlantic salmon, brown trout, or European eel.

Our results first demonstrate that archaeoichthyologists make few errors, and secondly, that we assign the species correctly for the majority of samples.



Map of samples used to test the pipeline on aDNA.

Thanks to our partners : BARBAZA Michel (TRACES), CLAVEL Benoît (AASPE), DACHARY Morgane (TRACES, SRA Nouv. Aquitaine), DAUJEARD Camille (HNHP), GEY Delphine (MNHN), GUILLAUD Émilie (AASPE), HAMILTON-DYER Sheila (Bournemouth Uni.), HARLAND Jennifer (Archeology Institute, UHI), HINGUANT Stéphan (INRAP), HORARD-HERBIN Marie-Pierre (CITERES), HUMMEL Esmee (Cultural Heritage Agency of the Netherlands), VAN DER JAGT Inge (ditto), JONSSON Leif (Gothenburg NHM), MADELAINE Stéphane (Musée National de la Préhistoire), MAKOWIECKI Daniel (Ins. Of Archeology, Poland), PERRIN Thomas (TRACES), PRIMAULT Jérôme (TRACES, ArScAn), QUINLAN Liz (Uni. Of York), SCHMÖLCKE Ulrich (Munich State Zoological Collection)