Inferring genomic signatures of cod anthropogenic evolution using ancient DNA and deep learning

Introduction: As populations face rapid environmental change, they must move, adapt, or perish. Understanding the genomic basis and rate of this rapid evolutionary change is a central goal in biology. Human-harvested species are of particular interest as their economic and social value through time has resulted in archaeological artifacts one can leverage to study the signatures of rapid evolution.[1] Here, I propose to investigate the genomic signature of rapid evolution in Atlantic cod (*Gadus morhua*), an economically and ecologically important species that has experienced rapid anthropogenic change in fisheries. Fisheries-induced evolution (FIE) has been demonstrated in cod phenotypes (e.g., body size, life history), but debate remains regarding the genomic mechanisms underlying this shift.[2,3] Additionally, because cod are often present in archaeological sites across the Atlantic, they present a rare opportunity for using ancient DNA (aDNA) time series data spanning roughly 1,000 years.[4] However, while aDNA data present powerful tools for studying populations through time, current analytical methods to test for signatures of selection using aDNA are limited, with current methods lacking the ability to detect signals from small effect alleles.[1,2,3] **Here, I propose to expand analytical methods for identifying selection from aDNA and apply those methods to test for genomic signatures of rapid evolution associated with the onset of fisheries in Atlantic cod.** Specifically, I will (*Aim 1*) create an expanded time series of cod genomic data, leveraging available archaeological data and pre-established aDNA resources in cod, (*Aim 2*) develop a novel deep learning model to detect selection from aDNA, and (*Aim 3*) apply the model to test for genomic signatures of historical and contemporary selection in Atlantic cod, specifically asking (**Q1**) How much has polygenic selection shaped cod genomic change over the past millennium?, and (**Q2**) Is the rate of genomic change from selection in Atlantic cod over the past millennium correlated with known fluctuations in fishing intensity?

Research plan: Aim 1: Collect a genomic time series. By generating a time series of ancient genomic data for Atlantic cod pre-industrial fishing to today, we can investigate trends in cod population genomics and assess correlations between fishing regimes and genomic change. By leveraging the distinct fishing histories of two cod populations (European and North American) over the last 1,000 years, we can better infer potential mechanisms of evolution. I will build on ongoing archaeological work to inventory and locate archived cod remains from at least 20 excavation sites across North America and Europe. Seeking broad coverage across geography and centuries, I will sequence 150-300 specimens in Cornell's aDNA lab. I will use contextual and radiocarbon methods to date cod bones and infer samples as ancestral to a particular population.[4] I will also supplement public cod genomic data on *GenBank* with 50 new samples from each side of the Atlantic using affordable low-coverage whole genome sequencing. Successful completion of this aim will result in a public cod genomic time series dataset spanning a millennium.

Aim 2: Develop a deep learning model. Existing neural network methods for inferring selection from genomic data are increasingly powerful but cannot effectively handle aDNA samples or infer polygenic selection.[5,6] Drawing from existing model architectures, [6] I will develop a deep learning model that infers selective responses from time series genomic data of variable quality. I will include data weighting and masking in the model to account for variable sample quality, including missing SNPs and low sequencing coverage. To infer polygenic selection, I will add layers to the model architecture that (1) allow input from multiple populations and (2) mimic existing methods for untangling signatures of polygenic selection from genetic drift and gene flow by analyzing covariance in genomic changes.[7] The complete model will combine the strengths of multiple previous selection inference methods while properly accounting for low-quality data.[6,7] The model will also output uncertainty in final inferences to assess precision. I will train the model on simulations mimicking known population demography using *SLiM 4*. I will use leave-one-out cross-validation of model performance and perform a sensitivity analysis by introducing mild perturbations in the data and observing output changes. Successful completion of this aim will result in a stand-alone method for analyzing genomic selection using aDNA data.

Aim 3: Infer the genomic basis of cod adaptation and correlations with fishing intensity. First, I'll ask **Q1**) How much has polygenic selection shaped cod genomic change over the past millennium? Because cod have undergone recent shifts in polygenic traits, I hypothesize that **H1**: Polygenic selection accounts for

significantly more genomic change in cod than other evolutionary processes. I will use two existing statistical methods to investigate signatures of selection over time in our data[8] and use our deep learning model to compare the total genomic change explained by polygenic vs. large-effect selection and other processes. I will interpret a greater proportion of genomic change associated with polygenic selection than other evolutionary processes as evidence for its dominant role in recent cod genomic evolution. I will also compare the relative importance of selection, drift, and gene flow in shaping cod genomes by comparing the proportion of inferred sites of selection consistent with particular evolutionary processes. Next, I will ask **Q2**) Is the rate of genomic change from selection in Atlantic cod over the past millennium correlated with known fluctuations in fishing intensity? A correlation between fishing and adaptive responses in cod has been elusive.[2,3] Our data is uniquely suited to answer this question by comparing genomic change across populations where fishing regimes varied through time.[9,10] I hypothesize that **H2**: The average selection coefficient across selected cod genomic sites increases after major fishing intensity shifts. I will identify dates of known shifts in fishing regimes from historical records of colonization and the introduction of fishing regulations, which are publicly available,[9,10] and generate a time series of average selection coefficients in each cod population using my novel deep learning model. I will use intervention analysis to assess the impact of fishing regime changes on selection. Comparative analysis between populations will validate the fisheries-induced nature of selection shifts, as some regime changes should affect only one of the two populations. Correlations between fishing intensity changes and selection in appropriate populations would provide evidence for fishing-induced genomic changes in cod. Otherwise, climate change and community composition shifts may be primary drivers of selective responses.

Feasibility: The extensive archaeological specimen collections, established protocols, and available expertise associated with my project make this proposal highly feasible. Proposed co-advisor Dr. Therkildsen (Cornell) manages an aDNA lab that collects and processes cod aDNA from identified archaeological sites and proposed co-advisor Dr. Messer (Cornell) is a computational biologist working with genomic time series and simulations. Cornell's Computational Biology's powerful computing cluster and my background in computer science make model development and training feasible.

Significance and intellectual merit: While AI methods are powerful emerging tools for understanding population genetic structure, they're currently ill-equipped to work with genomic data of variable quality and are not strong at identifying polygenic adaptation.[5,6] Our model will fill both gaps, opening doors for future studies on the genomic basis of rapid evolution in non-model organisms. We will clarify critical aspects of fishery-related evolution, such as whether phenotypic shifts result from the genomic evolution of phenotypic plasticity and if genomic changes correlate with fisheries, influencing the reversibility of anthropogenic genomic alterations.[1,2] Additionally, we will provide insight into the broader evolutionary question of whether rapid adaptation primarily involves a few large-effect or many small-effect alleles.

Broader impacts: By understanding how species adapt to sudden selective pressures, we can better identify which populations and species are most vulnerable to disturbance. Leveraging Dr. Therkildsen's involvement in NOAA's cod working group, results will be disseminated to managers, guiding real-world conservation efforts. I will also publish a large public cod genomic time series, a clear pipeline, and interactive online tutorials for future studies using our method, increasing accessibility to researchers interested in affordable methods for inferring species resilience and responses to change.

For community engagement, I will design a 3D exhibit for the Museum of the Earth, which features rotating exhibits showcasing work by Cornell students and faculty. By adjusting harvesting limits, visitors will visualize impacts on species, including altered size and life history in fish and reduction in size or complete loss of elephant tusks.[11] They will also be able to visualize recovery after exploitation, with the speed of recovery determined by the genomic basis of changes. An optional exit survey will assess how successfully the exhibit encourages visitors to think critically about their impacts and consumption habits.

References: [1]Dehasque et al. 2020 *Evo. Let.*; [2]Therkildsen et al. 2019 *Evolution*; [3]Hutchings & Kuparinen 2021 *PNAS*; [4]Martínez-García et al. *Proceedings B*; [5]Torado et al. 2019 *BMC Bioinf.*; [6]Whitehouse & Schrider 2023 *Genetics*; [7]Buffalo & Coop 2020 *PNAS*; [8]Illingworth et al. 2021 *Mol. Biol. Evol.*; [9]Holm et al. 2018 *Quat. Res.*; [10]Kraak et al. 2013 *Marine Policy;* [11]Campbell-Staton et al. 2021 *Science*