Ancient DNA reveals the Arctic origin of Viking Age cod from Haithabu, Germany

Bastiaan Star1,3, Sanne Boessenkool6, Agata T. Gondek1, Elena A. Nikulina2, Anne Karin Hufthammer4, Christophe Pampoulie8, Halvor Knutsen7,8,9, Carl André9, Heidi M. Nistelberga, Jan Dierking9, Christoph Petereit9, Dirk Heinrich4, Kjetill S. Jakobsen6, Nils Chr. Stenseth6,9,10, Sissel Jentoft5,3, and James H. Barrett1

1Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Blindern, NO-0316 Oslo, Norway; 2Centre for Baltic and Scandinavian Archaeology, Schleswig-Holstein State Museums Foundation, 24837 Schleswig, Germany; 3Department of Natural History, The University Museum, University of Bergen, NO-5020 Bergen, Norway; 4Marine and Freshwater Research Institute, IS-150 Reykjavik, Iceland; 5Institute for Marine Research, NO-4817 Fleidevigen, Norway; 6Centre for Coastal Research, University of Agder, NO-4604 Kristiansand, Norway; 7Department of Marine Sciences-Tjärnö, University of Gothenburg, 452 96 Strömstad, Sweden; 8GEOMAR Helmholtz Centre for Ocean Research Kiel, D-24105 Kiel, Germany; 9Boeselstrasse 9, D-24937 Flensburg, Germany; and 10McDonnell Institute for Archaeological Research, University of Cambridge, Cambridge CB2 2ER, United Kingdom

Contributed by Nils Chr. Stenseth, July 3, 2017 (sent for review June 8, 2017; reviewed by Michael Bunce and Matthew J. Collins)

Knowledge of the range and chronology of historic trade and long-distance transport of natural resources is essential for determining the impacts of past human activities on marine environments. However, the specific biological sources of imported fauna are often difficult to identify, in particular if species have a wide spatial distribution and lack clear osteological or isotopic differentiation between populations. Here, we report that ancient fishbone remains, despite being porous, brittle, and light, provide an excellent source of endogenous DNA (15–46%) of sufficient quality for whole-genome reconstruction. By comparing ancient sequence data to that of modern specimens, we determine the biological origin of 15 Viking Age (800–1066 CE) and subsequent medieval (1066–1280 CE) Atlantic cod (Gadus morhua) specimens from excavation sites in Germany, Norway, and the United Kingdom. Archaeological context indicates that one of these sites was a fishing settlement for the procurement of local catches, whereas the other localities were centers of trade. Fish from the trade sites show a mixed ancestry and are statistically differentiated from local fish populations. Moreover, Viking Age samples from Haithabu, Germany, are traced back to the North East Arctic Atlantic cod population that has supported the Lofoten fisheries of Norway for centuries. Our results resolve a long-standing controversial hypothesis and indicate that the marine resources of the North Atlantic Ocean were used to sustain an international demand for protein as far back as the Viking Age.

The global trade of animal products is driven by an increasingly international economy that geographically separates consumer demands from their ecological footprint elsewhere (1). However, long-range trade has an extensive history (2). For instance, in medieval and postmedieval Europe, the onset of urbanized market economies has been linked to the growth of long-range trade by historical and archaeological evidence (3, 4). The exploitation of increasingly distant fish populations has proven to be one of the clearest demonstrations of this ecological globalization (5–9). Fisheries around the coastal regions of the Lofoten Archipelago in Norway have a particularly long history; in this region, the regular arrival of seasonal spawning aggregations of Atlantic cod—migrating southwards from the Arctic Barents Sea (10)—coincides with those climatic conditions ideal for the freezing-drying and long-term preservation of cod without the use of expensive salt. These unique conditions allowed the development of an extensive long-distance trade and fishery that for centuries provided a high-quality yet affordable source of protein with a long shelf-life to urban centers around southern North and Baltic Sea regions (12). Identifying the emergence and growth of such extensive fisheries is a fundamental step in the study of past impacts of human exploitation (13–16). However, in the centuries before systematic historical records, this crucially depends on the ability to determine the biological source of archaeological samples.

Ancient DNA (aDNA) methods provide unique opportunities to elucidate diversification events, genetic admixture, or migration routes (e.g. see ref. 17) and have been used to assign historic individuals to their most likely geographic origin (18–20). In particular, the genome-wide analysis of 1,000s or more single nucleotide polymorphisms (SNPs) allows the determination of an individual’s geographic origin with accuracy and precision, even when limited genetic differentiation exists among regions (20–22). Since low levels of genetic differentiation are typical for marine species with high dispersal capabilities like Atlantic cod (23, 24), such genome-wide approaches are essential to increase the assignment power of modern (22) and ancient samples of these species (25). So far, however, aDNA studies using genome-wide approaches are lacking for marine fishes.

Here, we compare whole genome data of 15 Viking Age and medieval Atlantic cod from five archaeological sites to those of...
168 modern specimens from six populations (Fig. 1). We aim to discover where the cod were caught by analyzing fish bones from these sites. For example, were the cod eaten in a Viking Age (800–1066 CE) town of the western Baltic (Haithabu) caught in local waters, in the nearby North Sea or in truly distant waters of Arctic Norway (12), the Northern Isles of Scotland (26), and/or Iceland (6)? All of these distant regions are known to have produced dried fish for export later in the Middle Ages (27).

To improve our ability to identify the source of these ancient samples, we exploit our knowledge of patterns of genomic variation found among modern Atlantic cod populations. Distinct genomic regions with elevated population differentiation have been identified for this species (22, 28–30). Several of these regions are colocalized into four Mbp-scale, polymorphic inversions with high linkage disequilibrium (31–33). These polymorphisms segregate with a distinct geographical distribution and have been associated with temperature clinotypes and ecotype-specific migratory behavior (31, 32, 34). Their divergence can be used to improve the traceability of individuals (22) and, given their large size (between ∼5 and ∼17 Mbp), the inversion state of ancient samples can be easily determined from low coverage sequencing data.

We specifically focus on fish bones from Haithabu (n = 5, dated 800–1066 CE) for several reasons. First, a unique ninth-century account records the voyage of a Viking chieftain and trader from Arctic Norway to Haithabu. While his cargo is not fully specified—except for walrus tusks to be gifted to Alfred the Great of England (35)—a mundane consignment of dried cod may not have merited historical record. Second, artifactual evidence has confirmed contact between Haithabu and Scandinavian settlements of the North Atlantic (36). Third, Haithabu (and neighboring Schleswig, which later replaced Haithabu) preceded Lübeck as the main focus of trade in the western Baltic region. Because Lübeck came to control much of Europe’s dried cod trade (37), it is possible that its predecessors were already engaged in this exchange—at dates for which the historical record is incomplete. Fourth, the cod finds from Haithabu occurred alongside species such as saithe (Pollachius virens), ling (Molva molva), and halibut (Hippoglossus hippoglossus), which is more consistent with fishing in the North Sea or North Atlantic than in the Kattegat or Baltic Sea (38–40). Finally, isotope analyses using bone collagen (8, 41) or bone carbonate (9) suggest that the Haithabu cod were not locally caught, although their origin remains ambiguous. This previous research pinpoints the fish bones from Haithabu as ideal material with which to test the hypothesis of Viking Age transport of cod from northern Norway by using genomic methods. Without definitive interpretation of the Haithabu cod, the earliest secure historical and archaeological evidence for dried cod imports to towns around the North and Baltic Seas dates to the 12th, 13th, and 14th centuries (7, 8, 37, 42, 43). Fishing in Arctic Norway occurred long before the Viking Age (44), but most arguments for pre-12th-century long-range exports have had to rely on historical sources such as Icelandic sagas that postdate the events that they nominally described (43, 45).

In addition to the specimens from Haithabu, we include two bones (dated 1100–1280 CE) from Schleswig, one bone (dated 700–950 CE) from the small inland trading settlement of Bjørkum in western Norway and two bones (dated 1025–1175 CE) from Oslo. Lastly, five cod bones (dated 1000–1200 CE) are from a fishing settlement in Orkney, northern Scotland, where a local catch can be confidently assumed (26).

Results

We obtained 1,013 million paired reads (average read length 40–78 bp) that contained 15–46% endogenous DNA and resulted in 1- to 3.4-fold nuclear coverage for 15 of 19 ancient cod samples (Table S1). Analyses of postmortem degradation patterns showed the typical fragmentation and elevated deamination rates as expected from authentic ancient DNA (Fig. S1). For the modern data (Fig. 1), an average of 47 million paired reads were obtained per specimen, resulting in ∼ninefold individual coverage of the nuclear genome (Table S2). After SNP calling and filtering, we obtained a dataset of 156,695 SNPs.

First, we compared the genome-wide diversity of ancient samples and modern samples using principal component (PCA) and ADMIXTURE analyses. We excluded the four chromosomes [linkage group (LG)01, 02, 07, and 12] that contain the large inversions (31–33), as these distort genetic analyses that assume linkage equilibrium. The PCA showed that the three modern populations—North East Arctic (NEA), Lofoten and North Sea—form a large, overlapping group, with Iceland clustering near the NEA (Fig. 2A). The majority of the ancient samples cluster with this large group. The modern eastern Baltic and Øresund form two distinct groups, with one ancient specimen from Schleswig clustering with the Øresund. A lack of strong genetic population structure between most modern samples is supported by ADMIXTURE, for which the best fit [based on lowest cross-validation (CV) error] is two populations (k = 2). Here, Iceland and eastern Baltic have distinct ancestry, whereas the other modern populations show a more mixed ancestry (Fig. 2B). The Øresund has a higher level of Baltic ancestry than the other admixed populations. All ancient individuals have mixed ancestry, with the one Schleswig sample having a similar level of Baltic ancestry to the modern Øresund. At a lower optimal fit, ADMIXTURE estimates biologically plausible ancestry for k = 3 (identifying an Øresund component) and k = 4 (identifying a North Sea component). Nonetheless, for k = 5, the model separates the original eastern Baltic cluster rather than differentiating either NEA or Lofoten cod (Fig. S2). Overall, these analyses show that the ancient individuals are genetically most related to four modern populations: one Schleswig individual is genetically more similar to the Øresund and all other individuals clearly classify to the NEA, Lofoten, or North Sea population.

To further differentiate the origins of the ancient individuals, we determined their genotypes for the inversion loci on LG01, 02, 07, and 12. Modern populations can have divergent allele
frequencies for these loci (31, 32), and the NEA population is set apart by having inverted alleles near fixation at all four loci (Fig. 3A). PCAs of these inversions for the four modern populations show the typical trimodal clustering of a biallelic locus with heterozygote genotypes clustering intermediate to the two divergent homozygotes (46) (Fig. S3). All ancient individuals cluster within this trimodel pattern, allowing for the decisive determination of their genotypes at each inversion locus (Fig. S3). Collinear alleles dominate in the Orkney samples, while inverted alleles dominate in Haithabu (Fig. 3B). The Bjørkum sample has a predominantly inverted composite genotype, whereas Schleswig and Oslo show a mixed pattern, with one specimen having a collinear and the other having a more inverted composite genotype, in each location, respectively.

Based on their inversion genotypes, we investigated if ancient individuals statistically differed in their resemblance toward a particular modern population. The probability of obtaining an inversion genotype follows a binomial distribution given the underlying allele frequency in a population. We can assume independence between loci since the inversions are located on different chromosomes. It is thus straightforward to calculate the overall probability of obtaining a composite ancient inversion genotype—based on the four modern populations’ respective allele frequencies—as a measure of an individual’s affinity toward a specific population. Based on this probability, we find that, apart from a single Haithabu specimen, the ancient samples show two types of affinity (Fig. 3C); individuals with a predominantly collinear composite genotype have a >99% probability of being drawn from the Lofoten, the North Sea, or the Øresund population. Conversely, individuals with a predominantly inverted composite genotype have a >99% probability of being drawn solely from the NEA population. The atypical Haithabu specimen has a >99% probability of coming from either the NEA or Lofoten population.

Discussion

By investigating genome-wide patterns of variation—including four megabase-scale inversions (31–34)—we show that ancient cod specimens from Viking Age (Haithabu and Bjørkum) and early medieval (Schleswig and Oslo) trading sites have a genomic affinity to the modern North East Arctic population. This finding has archaeological and evolutionary implications.

First, this study provides a unique genome-wide study of archaeological fish bone and demonstrates its potential as an archive of ancient DNA. We obtained short reads (<100 bp), with the typical fragmentation and cysteine deamination patterns expected after postmortem degradation (47–49). These results are therefore fully consistent with the extraction and analysis of authentic ancient Atlantic cod DNA. Not all types of animal bone preserve DNA equally well, and significantly higher proportions of endogenous DNA have been found in dense and heavy bone types such as the petrous bone (50–52). It is presumed that it is the high density of the bone that leads to reduced

Fig. 2. Genetic population structure in 183 Atlantic cod specimens. (A) PCA based on 99,819 SNPs. Ancient specimens (stars) were projected onto the first two principal components calculated by using individuals from modern populations (circles). (B) ADMIXTURE ancestry components (k = 2) for modern and ancient specimens. The width of the bar for ancient specimens is widened to aid visualization. LG01, 02, 07, 12, and unplaced scaffolds were excluded from these analyses (see Results for explanation).
bacterial and chemical-mediated decay and improved DNA preservation (50, 53). This hypothesis would suggest that fish bones—that are porous, brittle and light—should be a poor source of DNA. Instead, we find surprisingly high levels (15–46%) of endogenous DNA preservation in 15 of 19 fish bone specimens up to 1,300 y old from five different archeological sites. Notwithstanding the use of a recently developed extraction protocol aimed to maximize endogenous DNA (54), this rate of success compares favorably to results from mammalian bones whereby the majority of samples—excluding petrous bones—typically yield a few percent endogenous DNA at most (55). Our positive results agree with studies using PCR-based methods that have reported successful amplification from fish bones (56–60) in some cases up to 10,000 y old (61). The observation that porous, light fish bones can yield whole genome shotgun libraries with high levels of endogenous DNA underscores our lack of understanding of DNA preservation in different types of animal bone. These results also illuminate the potential of the large reservoir of archaeological fish bone as a source for aDNA of sufficient quantity and quality to study long-term evolutionary processes in the marine environment.

Second, we identify polymorphic chromosomal inversions in ancient Atlantic cod specimens. Chromosomal inversions are expected to play a major role in ecological adaptation (62, 63). In cod, these regions contribute to elevated genomic diversification between modern populations and ecotypes—despite low overall levels of divergence (25, 31–34, 64)—and they have been suspected to be under selection (22, 28, 29). The high genomic divergence of these alleles, in combination with their wide geographic distribution, suggests that these have been maintained as a polymorphism for hundreds of thousands, if not millions of years (32, 33). Here, we directly observe the millennium-long maintenance of such inversions, which provides an opportunity to investigate their temporal stability. Although speculative due to the currently small samples sizes from Orkney (n = 5) and Haithabu (n = 5), our results suggest that the divergent frequency distributions of these inversions indeed have remained stable in their respective populations. Given the low overall genomic differentiation outside of these inversions in modern populations (31–33), it appears that this divergence is sustained despite ongoing gene flow, which further supports a hypothesis that these alleles are maintained as a balanced polymorphism (65).

Finally, we solve a long-standing hypothesis generated by history, archaeology, and stable isotope analysis that dried cod from northern Norway were transported during the Viking Age to Haithabu. Currently, the NEA cod feeds in the Barents Sea and its spawning grounds are restricted to the northern coasts of Norway, especially (although not exclusively) in the area around Lofoten (10, 11). One could argue that this population spawned in the Skagerrak or Kattegat during the Viking Age, but such a scenario is implausible; historical records of the Norwegian fisheries since the 12th century show that spatial fluctuations in the distribution of fishing effort targeting these spawning aggregations have been restricted (43, 66). Furthermore, our ancient data are also consistent with observed long-term spawning fidelity in the North Sea region (67, 68); the genetic affinity of the ancient Orkney population, a focus of fishing since the Viking Age (69), agrees with those of modern individuals living in the North Sea. Similarly, we observe the specific affinity of one ancient Schleswig specimen toward a distinct population in the Øresund (70). This observation also suggests local spawning fidelity and agrees with fine-scale population structure reported for this region (70–73). Overall, by identifying their most likely source population, we conclude that the ancient Haithabu cod were not caught locally, but around the Norwegian coast during historic spawning aggregations of NEA cod. Since climatic conditions restrict the production of dried cod without salt to the north of Norway (66), we can further constrain their source to this northern region, implying transportation over long spatial distances during the Viking Age. Salting of cod did not begin in Norway until the 1690s, when klipfish was introduced to compete with new international products from locations such as Newfoundland. Before that, Norwegian farmers, Sami hunters, and sometimes immigrant fishermen instead produced air-dried stockfish without salt, a practice limited to the north, especially Lofoten (43, 74). We cannot yet infer conclusions about the scale of cod transport and/or trade in the Viking Age, potentially ranging from providing travelers’ rations to supplying an urban staple, which—together with obtaining a more refined chronology—should be a subject of future research. Moreover, knowing the Haithabu results, it becomes important to ask whether cod bones from even earlier trading sites in the

![Spatial genomic variation in megabase-scale inversions in Atlantic cod.](image-url)
western Baltic region (e.g., Groß Strömendorf) might also represent fish from a distant source (75).

Conclusion

Our discovery of distinctive genomic inversions in ancient cod specimens has made it possible to answer the long-standing question of whether dried cod was transported from northern Norway during the Viking Age (800–1050 CE), solving a mystery epitomized by the ninth-century account of an Arctic Norwegian chieftain’s voyage to Haithabu. Our findings suggest that distant requirements for Arctic protein had thus already begun to influence the economy and ecology of the north over the chronology under consideration. Our study highlights the potential of coupling modern genomics with ancient DNA to study the origins of historic trade routes in fish and other taxa.

Materials and Methods

The ancient samples (n = 19) are from the archives of excavations conducted at Haithabu (39), Schleswig (76), Bjerkum (77), Oslo (78), and Orkney (26) and are dated based on archaeological context. They come from waterlogged (Haithabu, Schleswig, and Oslo) and free-draining (Bjerkum and Orkney) deposits and have been stored dry and unfrozen in fluctuating ambient temperatures after excavation. Bones were morphologically identified as Atlantic cod and selected from different archaeological layers, from fish of differing size and/or from the same element to avoid multiple samples from individual fish (Table 51).

Extraction and Library Creation. DNA from ancient samples was extracted in a dedicated aDNA laboratory at the University of Oslo following strict precautions (79, 80) using a combined bleach and predigestion (BleDD) protocol (54). Ancient DNA libraries were created by using a blunt-end ligation protocol (81) with minor adjustments (20) (SI Materials and Methods). DNA from modern samples (n = 168, Fig. 1 and Table 52) was extracted—individually in a separate laboratory from the ancient samples—using a DNeasy Blood & Tissue Kit (Qiagen) and sheared to an approximate insert size of 350 bp (82). Modern libraries were created by using a TruSeq DNA PCR-Free Preparation Kit. All libraries were sequenced on an Illumina HiSeq 2500.

Data Processing. The ancient read data were processed by using PALEOMIX (83). In short, forward and reverse reads were collapsed with AdapterRemoval v1.5 (84) and aligned to the GadMor2 reference (85, 86) by using BWA v.0.7.5a-r405 (87). The modern data were aligned by using BWA mem. Reads not aligned with the minimum quality score (MapQ) of 25 used for subsequent analyses. aDNA damage patterns were investigated by using mapDamage v.2.0.6 (49).

SNP genotypes were obtained by using GATK v. 3.4.46 (88), with duplicate removal (Picard Tools v. 1.96) and indel realignment (GATKs IndelRealigner). Genotypes were jointly (GATKs Genotypecaller) called for modern and ancient samples separately with default settings, allowing a maximum of three alternate alleles. The modern data were filtered with BCFTOOLS v. 1.3 (89) by using filter e: “FS>60.0 || MQRankSum<12.5 || ReadPosRankSum<8.0 || QD<2.0 || MQ<40-“SnvGap 10” and VCFTOOLS v.0.1.14 (90), keeping biallelic loci with a maximum average read depth of 30 and a minimum MAF of 0.05. The filtered modern and ancient dataset were binned as BCFTOOLS lice, after which genotypes with a quality below 15 and read depth below 3 were set as missing and all C > T and G > A SNPs were removed. The final dataset consisted of 156,695 SNPs.

Analyses. For inferring genome-wide population structure, SNPS were pruned (~indep-pairwise 100 10 0.5) for linkage disequilibrium (LD) by using PLINK v1.90p (91) and LG01, 02, 07, and 12 were excluded. PCA was performed with smartPCA, EIGENSOFT v.6.1.4 (92) whereby ancient individuals were "projected" by using "lsqproject" to account for missing data. Model-based clustering was performed using ADMIXTURE v1.3 (93). PCA plots for the inverted regions (LG01, 9.1–26.2 Mbp; LG02, 18.5–24 Mbp; LG07, 13.6–23 Mbp; LG12, 13.1–13.6 Mbp) were generated without LD pruning (46). Finally, the probability of obtaining the ancient individual’s composite inversion genotype from the allele frequency distribution of the four modern populations was calculated by binomial sampling of genotypes and scaling these probabilities to one.

Acknowledgments. We thank M. Skage, S. Kollas, A. Tooming-Klunderud, and H. Rydebeck from the Norwegian Sequencing Centre for sequencing and processing of samples; and M. H. Hansen for extracting modern Atlantic cod specimens. P. R. Berg, M. Malmström, and K. Berndt assisted in sampling modern cod from Norway and the Øresund, respectively. U. Schmölke kindly facilitated sampling at the Centre for Baltic and Scandinavian Archaeology; and M. Ramstad of the University of Bergen provided access to the Bjerkum assemblage. This work was supported by Research Council of Norway Projects “Tracking Viking-assisted dispersal using ancient DNA” (230821/F20), “Fisheries induced evolution in Atlantic cod investigated by ancient and historic samples” (203850/E40), “The Aqua Genome Project” (221734/O30), and Leverhulme Trust Project “Northern Journeys: Reimagining the Medieval Revolution and its Aftermath” (MRF-2013-065).