

This is a repository copy of *Barcoding the largest animals on earth:on-going challenges and molecular solutions in the taxonomic identification of ancient cetaceans*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/103734/>

Version: Accepted Version

Article:

Speller, Camilla Filomena orcid.org/0000-0001-7128-9903, van den Hurk, Youri, Charpentier, Anne et al. (8 more authors) (2016) Barcoding the largest animals on earth:on-going challenges and molecular solutions in the taxonomic identification of ancient cetaceans. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 20150332. ISSN 1471-2970

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Barcoding the largest animals on earth: on-going challenges and molecular solutions in the taxonomic identification of ancient cetaceans

Camilla Speller^{*a}, Youri van den Hurk^b, Anne Charpentier^c, Ana Rodrigues^c, Armelle Gardeisen^d, Barbara Wilkens^e, Krista McGrath^a, Keri Rowsell^a, Luke Spindler^a, Matthew Collins^a, Michael Hofreiter^{f*}

^a BioArCh, Department of Archaeology, University of York, Environment Building, York, North Yorkshire, YO10 5DD, UK orcid.org/0000-0001-7128-9903

^b Institute of Archaeology, University College London, 31-34 Gordon Square, London, WC1H 0PY, UK

^c CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE – CNRS, Montpellier Cedex 5, France

^d Archéologie des Sociétés Méditerranéennes, UMR 5140, CNRS, Labex Archimède IA-ANR-11-LABX-0032-01, Université Paul-Valéry Montpellier, F-34970 Latte

^e Dipartimento di Scienze della Natura e del Territorio, Università degli Studi, Sassari, Italy

^f Institute of Biochemistry and Biology, Faculty of Mathematics and Natural Sciences, University of Potsdam, 14476 Potsdam, Germany

Keywords:

Ancient DNA; Archaeozoology; Cetaceans; Collagen peptide mass fingerprinting; Species identification; ZooMS

*Author for correspondence (camilla.speller@york.ac.uk; michael.hofreiter@uni-potsdam.de).

Main Text

Summary

Over the last few centuries, many cetacean species have witnessed dramatic global declines due to industrial overharvesting and other anthropogenic influences, and thus are a key target for conservation. Whale bones recovered from archaeological and paleontological contexts can provide essential baseline information on the past geographic distribution and abundance of species required for developing informed conservation policies. Here we review the challenges with identifying whale bones through traditional anatomical methods, as well as the opportunities provided by new molecular analyses. Through a case study focused on the North Sea, we demonstrate how the utility of this (pre)historic data is currently limited by a lack of accurate taxonomic information for the majority of ancient cetacean remains. We discuss current opportunities presented by molecular identification methods such as DNA barcoding and collagen peptide mass fingerprinting (ZooMS), and highlight the importance of molecular identifications in assessing ancient species distributions through a case study focused on the Mediterranean. We conclude by considering high-throughput molecular approaches such as hybridisation capture followed by next-generation-sequencing as cost-effective approaches for enhancing the ecological informativeness of these ancient sample sets.

1. Introduction

Humans have been exploiting cetaceans for thousands of years, first through the opportunistic use of stranded or drift whale carcasses, and subsequently by active hunting [1–4]. Their value came from the use of meat and blubber as food, blubber as fuel in oil-burning lamps, teeth (of odontocetes) as a valuable form of ivory, baleen (of mysticetes) as a raw material source, and bones used for building purposes, tool production and as solid fuel (given their high oil content) [1,5–8]. Intensive human exploitation (particularly the industrial hunting practices of the 19th and early 20th century), as well as other anthropogenic influences (e.g. sonar, ship strikes, habitat degradation, etc.) reduced the size of whale populations worldwide and even extirpated some local populations [9,10], including the eastern North Atlantic populations of right whale (*Eubalaena glacialis*) [11] and the Atlantic populations of gray whale (*Eschrichtius robustus*) [12]. The past few decades have witnessed major efforts in the conservation of whales, including the 1984 moratorium on commercial whaling instituted by the International Whaling Commission (IWC), and the Convention on the International Trade of Endangered Species (CITES). Although many whale species are now protected and some populations are recovering [9,13], cetacean conservation is still a key ecological priority for many countries. Developing informed conservation policies and sustainable management plans require accurate historic data on cetacean abundance and distribution at various stages in their interactions with humans. Archaeological and paleontological records are key to the reconstruction of these ecological baselines [14], but they have been dramatically underused, largely because of the challenges

associated with the taxonomic identification of ancient whale bones. Molecular methods have advanced substantially in the past few decades, and molecular barcoding now provides a new opportunity to decipher and maximise the information potential of the archaeological and paleontological records.

In this paper, we review the challenges with identifying ancient whale bones, as well as the opportunities provided by new molecular identification methods. We begin by summarising the limitations inherent to taxonomic identification based on traditional anatomical methods, illustrated with a case study from the North Sea on the proportion of unidentified archaeological cetacean remains housed in museums and repositories. We then discuss the opportunities for more accurate identifications made possible by molecular analyses, and demonstrate the need for molecular validation through a case study comparing anatomical and molecular identifications of whale bones from Mediterranean archaeological contexts. Finally, we conclude by presenting future perspectives for molecular methods, including high-throughput approaches for the study of ancient cetacean assemblages.

2. Limitations in identifying whale bones using anatomical methods

In spite of the millennia of human-cetacean interactions, the research potentials of paleontological or archaeozoological cetaceans have received very little attention, in large part because of the difficulty in identifying (often fragmentary) ancient whale bones to the genus or species level using comparative anatomy methods. Compared to other large mammals, whale bone is extremely friable; composed primarily of oil-filled cancellous bone, with only a thin external cortical layer, whale bone easily breaks up into non-diagnostic fragments. When a whale is exploited through active hunting or scavenging of drift carcasses, their sheer size limits the viability for humans to transport complete anatomical elements far from the beach [15]. Thus, in archaeological contexts, the larger the animal, the less bone is transported from shore to settlement, decreasing the likelihood of finding diagnostic pieces of the skeleton. A single animal can also supply up over 40 metric tons of bone [16,17], making it difficult to distinguish the number of species or individuals represented by fragmentary remains. The use of cetacean bone as raw material for combustion or tool production further fragments and modifies the bone [5]. Even when diagnostic elements are preserved, the range of morphological variation present among and within (e.g. sexual dimorphism) species can confound taxonomic identifications [18].

These identification problems are compounded by a lack of comprehensive or easily accessible skeletal reference collections, which are usually restricted to a few large national natural history museums (e.g. National History Museum, London, England or Naturalis in Leiden, the Netherlands) [19]. Unlike most other mammalian collections, the range of morphological variation present within each species is not well represented, and is thus not well characterized in taxonomic identification atlases for a wide diversity of bones. Indeed, the challenges with storing such huge specimens mean that repositories do not typically curate more than one or two individuals from each species, often only retaining particularly diagnostic elements, such as the cranium. Collections are particularly incomplete for populations that were extirpated prior to the creation of modern museum collections (from the 18th century), such as the North Atlantic right whale (functionally extinct in the eastern North Atlantic [11]) or the Atlantic population of the

gray whale (extinct [12]). Even the most complete collections may not serve as representative guides for ancient remains, as archaeological specimens may be considerably larger than museum specimens curated relatively recently, due to the diminution in the overall size of mature animals following the advent of modern whaling [18].

These difficulties create substantial gaps and biases in the archaeological record, with ramifications for understanding past human interactions with these marine mammals. For example, in a study of whale remains in the Western Isles (north west of Scotland), Mulville [5] noted an increase in both the proportion and taxonomic diversity of whale bones from the Later Bronze age through to the Norse age, with an increase in the proportion of large whale species. However, whilst progressively more species were identified through time, an increasing proportion of remains were not taxonomically identifiable; only 30 of 568 examined whale bones could be identified to species, largely due to extensive modification or burning of the bones [5]. Similarly, in a collection of fifty archaeological specimens from seven North Atlantic archaeological sites, ranging from the Mesolithic until the Early Modern period, most of the bone fragments could be morphologically identified only to 'marine mammal' or 'cetacean' [20]. Finally, a study in the Northeast Pacific Coast of North America found that although whale bones were recovered from Nuu-chah-nulth (Nootka) sites as early as 4000 BP (before present), fewer than 20% of these could be identified to species [21].

(a) Case study: ancient cetacean assemblages in the North Sea

A case study from the southern North Sea (Figure 1) illustrates the difficulties with the identification of cetacean species in the zooarchaeological record. This synthesis of faunal data from published archaeological reports revealed at least 102 sites with preserved cetacean remains, the majority of which date to the early Medieval period [22]. Of the 616 remains recovered, less than half ($n=306$) could be morphologically identified to the species level through traditional comparative anatomy methods (Supplementary Table 1). Furthermore, among these taxonomically identified bones 119 originate from a single site, the early medieval site of Flixborough (represented by 115 common bottlenose dolphin *Tursiops truncatus*, 3 minke whale *Balaenoptera acutorostrata*, and 1 killer whale *Orcinus orca*) [23].

Overall, most of the identified specimens across all 102 sites represented dolphins or porpoises. The taxonomic identification of these small species is significantly easier, as a greater proportion of entire bones are preserved and more complete reference collections are available. Only 12 specimens (<4% of the identified assemblage) were identified as baleen whale species. The fact that half of these baleen whales were identified as North Atlantic right whale demonstrates the significance of these ancient sample sets. Indeed, this species has all but disappeared from the North Sea, but its prevalence in these archaeological remains (albeit within an extremely small sample set) hints at its potential historic abundance within the region. It is highly likely that a significant proportion of the unidentified cetacean specimens are also from large (baleen) whales, as they are less likely to be taxonomically identified if they are in a fragmented state compared to smaller species (see discussion in [20]). Human behaviour may also preferentially increase fragmentation of large species compared to their smaller counterparts. Compared to dolphins or porpoises, baleen whales provide a more abundant supply of bone, with thicker cortex, making them better suited as a raw material for tool production [24,25]. The lipid content of large whales is also higher than that of dolphins [26] making them more desirable as sources of biofuels. The deliberate fragmentation of these oil rich elements to liberate the oil or

maximize the surface area for burning decreases the likelihood of morphological identification [16,27]. In the North Sea, and in other regions, this lack of taxonomic precision limits our ability to detect historic changes in cetacean distribution and abundance, and document how these populations have been impacted by human activities.

3. Opportunities from molecular identification techniques

(a)DNA barcoding

Over the last two decades, molecular methods have been increasingly applied to the problem of cetacean identification, but their primary focus was the study of contemporary populations. Indeed, given the 1984 moratorium on commercial whaling by the IWC, their protected status under CITES, and the many national laws protecting particular species and populations, accurate taxonomic identification of whale products has become essential to differentiate products obtained from legal versus illegal exploitation or trade. For example, identification to the species or even population level may be key to assessing whether whale products (skin, blubber, meat) sold in domestic markets have a legal origin (e.g., if they come from small odontocetes not covered by the IWC moratorium, or from populations exploited under aboriginal subsistence permits) or not, but such products are often processed in ways that render morphological identification impossible. Considering that such processing (cooking, salting, drying, marinating) may significantly degrade DNA, early molecular studies targeted mitochondrial DNA (mtDNA), amplifying relatively short diagnostic fragments (150-500bp) of the control region or cytochrome b (cytb) gene to identify taxa and to estimate geographic provenience [28,29] (Figure 2a). By comparing the resulting sequences to a databank of known species and populations, it is possible to evaluate the relationships between known and unknown samples by parsimony or maximum likelihood criteria, with the reliability of phylogenetic relationships analyzed by bootstrapping procedures [30]. Tree-based approaches, however, can be problematic in situations where relationships among interbreeding organisms are not hierarchical, or where species are polyphyletic [31]. In these cases, taxonomic identifications may be more robust when sequences are analyzed using vector or distance-based clustering methods to evaluate distribution of derived character states [31,32].

In the 1990s, genetic databases were limited in the number of type species and populations represented, and thus taxonomic identifications and/or phylogeographic analyses were often tentative [29]. Over the last two decades, follow-on studies in cetacean systematics and phylogeography [33,34], the development of comprehensive wildlife DNA registers [35] and validated reference sequence databanks (e.g. DNA surveillance [36]) have significantly enhanced the ability to 'barcode' morphologically ambiguous cetacean remains, not only in markets but also animals caught as fisheries bycatch or derived from strandings. Beyond species identification, mtDNA and nuclear DNA (short tandem repeats (STRs), actin sequences) are being applied to quantify the minimum number of individuals entering trade [37], estimate the total catches resulting from market meat[38], or even to track the life-history of an individual whale [39].

Studies have also explored the potential for applying these molecular methods to paleontological or archeological remains. As with modern whale product identification, studies of ancient

specimens primarily concentrated on recovering short diagnostic fragments of mtDNA control region [40,41], cytb gene [42] or both [43,44] for accurate taxonomic identification. In addition to archaeological bone, DNA analysis has been applied to other whale products, such as museum samples of baleen [45,46], whale ivory or scrimshaw [40], with relative success. Although there has been the occasional large-scale study identifying hundreds of samples [47], DNA-based studies have been primarily applied to demonstrate the feasibility of these molecular techniques, or at the site level, to identify the range of species exploited within a geographically restricted region. Among the aforementioned specimens in the North Sea case study, only two have been identified through ancient DNA analysis (two fin whale specimens, at Barreau Saint Georges, France) [48]. There is thus much unexploited potential for the application of these methods to the analysis of ancient specimens. However, the relatively high cost of these analyses and the fact that they need to be done in specialised laboratories (to prevent DNA contamination) remains a limiting factor to their large-scale application.

(b) ZooMS: Collagen Peptide Mass Fingerprinting

Peptide mass fingerprinting (PMF) has been widely used as a rapid and cost-effective protein identification method based upon the pattern of mass to charge (m/z) ratios [49]. Most recently, it has been developed for the most abundant protein in archaeological bone: collagen (Figure 2b). In mammals, collagen is composed of two alpha 1 chains, and a third, more rapidly evolving alpha 2 chain. In collagen PMF approaches, collagen is extracted from archaeological bone, followed by enzymatic digestion which cleaves proteins at specific amino acid sites producing a characteristic mixture of peptides. The peptides are analyzed through Matrix-assisted Laser Desorption/Ionisation Mass Spectrometry (MALDI-MS), producing a 'peptide mass fingerprint' based on their respective m/z ratios. Species identification of archaeological bones can thus be accomplished by comparing collagen peptide fingerprints with the fingerprints from known samples - i.e. Zooarchaeology by Mass spectrometry (ZooMS) [50,51]. Collagen's relatively slow rate of evolution means that it is variable enough to discriminate between mammal genera, but is sufficiently similar to map differences across broad taxonomic groups, such as cetaceans [52,53]. The ZooMS approach has been developed and tested on North Atlantic cetacean species, providing a rapid and cost-effective identification screening approach often to the genus or species level [20,54].

The advantages of a collagen versus DNA-based approach for identifying ancient samples are numerous. First, collagen is a remarkably robust protein, and recent evidence suggests that collagen survives at least 10 times longer than DNA, preserving even in tropical climates where DNA preservation is poor [53,55]. Unlike PCR-based approaches, which can be limited by primer specificity, ZooMS can be applied to highly fragmented non-diagnostic bone without any prior taxonomic knowledge [56]. Since ZooMS does not require the amplification of degraded ancient molecules, the risk of false positives from contaminating modern template or previously amplified PCR products is also reduced. Moreover, collagen can be recovered and analyzed using a non-destructive ammonium bicarbonate buffer which enables bone samples or artifacts to be analyzed without destructive sampling [57]. ZooMS, however, does have its limitations: due to the relatively slow mutation rate of collagen, taxonomic precision is often limited to the genus level. For example, although most baleen whale species can be distinguished, ZooMS cannot currently differentiate between bowhead and right whale, or among some dolphin species [20].

Additionally, robust identifications often require the successful recovery of multiple diagnostic peptides. Thus, mass spectra from poorly preserved samples may only allow identification to higher taxonomic levels (family, order) if diagnostic peptide markers are absent. Genetic methods may be required to clarify species identity, and are certainly required for identification to the subspecies or population levels. However, applying ZooMS as an initial screening method can provide a cost-effective preliminary identification, as well as insight into overall biomolecular preservation and the likely success for subsequent DNA analysis [58] or radiocarbon dating [59].

Biomolecular identification approaches such as DNA barcoding and ZooMS can offer robust taxonomic identifications of ancient cetacean, however, they can be limited by taphonomic histories and biomolecular preservation. Some studies, for example, have noted a high presence of inhibitory substances in ancient whale bones, compromising the success of PCR amplifications [60–62]. Also, archaeological whale bone has often been burned, limiting the quantity and quality of DNA and collagen that can be obtained from the samples [63,64]. Likewise, biomolecular degradation can be extensive in samples recovered from tropical or subtropical environments [65,66]. In spite of these challenges, biomolecular identifications of archaeological cetaceans can be applied to many specimens that remain currently unidentified, and thus make a decisive contribution to reconstructing the ecology and population history of cetacean species.

(c) Importance of molecular identifications: a case study validating ancient cetacean specimens in the Mediterranean Sea

In order to illustrate the necessity of validating osteological identifications of cetaceans, we present a case study of 17 pre-industrial cetacean specimens from six sites around the Mediterranean Sea (Table 1, Figure 3, Supplementary Table 2). This collection is particularly meaningful as it includes five specimens previously identified through comparative anatomy methods as Atlantic gray whale (*E. robustus*) [67]. The gray whale is currently found only in the North Pacific, and the circumstances of its disappearance from the North Atlantic remain a mystery [12] as this population left very few historical, archaeological, or paleontological traces. Fewer than 60 remains are known from both sides of the Atlantic, and the 34 records in the eastern North Atlantic (dated from the late Pleistocene to the 18th century), are nearly all from the North Sea [68] (Figure 3). The restricted spatial distribution of these bones is likely a poor reflection of their actual past range; indeed, habitat modeling predicts gray whale would have also occurred further south, including the Bay of Biscay, and to lesser extent, the Mediterranean Sea [68]. Due to this paucity of remains, the reliability of each new gray whale identification outside the currently known distribution is potentially crucial to our understanding of the distribution and ecology of this population. Twelve additional Mediterranean samples (identified only to the level of cetacea), were also included in this study to further increase the possibility of detecting gray whale remains.

We analysed the 17 ancient bones using DNA barcoding (cytb mtDNA analysis) and ZooMS (methods described in Supplementary Materials), identifying 11 fin whale (*Balaenoptera physalus*), one sperm whale (*Physeter catodon*), one right whale (*Eubalaena glacialis*), and one Cuvier's beaked whale (*Ziphius cavirostris*). The relative merits of both techniques in terms of their precision and robusticity are exemplified here. For example, although ZooMS identified

two samples only as ‘beaked whale’ and ‘bowhead/right whale’, respectively, DNA provided the resolution to confirm these as Cuvier’s beaked whale and right whale. Although ZooMS may be less precise, it may be more successful with poorly preserved samples: three samples that failed DNA analysis were identified through ZooMS as two fin whales and one baleen whale (Mysticeti), respectively. For the latter sample, a higher taxonomic resolution was not possible due to a lack of high molecular weight diagnostic peptide markers in the mass spectra (Figure S1). Despite the advantages of applying both techniques to the same assemblage, two samples failed to produce identification using either molecular method, illustrating the limitations of working with degraded archaeological materials.

Of the five samples previously believed to correspond to gray whale, three were identified as fin whale and one as sperm whale, with the fifth identified only as a ‘baleen whale’ (although until further analyses are conducted, the gray whale identification of this latter specimen should be considered unreliable). These results illustrate the necessity of validating identifications based on anatomical identifications through molecular techniques, even more so as this is not the first study to reveal previously misidentified whale remains. For example, a mtDNA-based analysis of 17th century Basque whaling remains determined that morphological identifications previously assigned to right whale were in fact bowhead whale [69]. Likewise, a molecular analysis of archaeological cetacean fragments from Tierra del Fuego believed to correspond to the remains of a single animal within a hunter-gatherer midden revealed the presence of multiple whale species as well as non-cetaceans (e.g. human, pinniped) [42]. This latter study also demonstrated that available museum reference specimens may themselves have been incorrectly identified using anatomical methods. These and other studies (e.g. [47,70]) collectively highlight the crucial need for the molecular screening of existing and future zooarchaeological collections containing whale bones.

The hypothesis that gray whales previously migrated to calving grounds in the Mediterranean sea was largely supported by the presence of these five ‘putative’ gray whale bones [66] - identification which failed to be confirmed by molecular methods in this study. However, records for fin, sperm and beaked whales are in agreement with the composition of the Mediterranean whale assemblage: fin whales are the most common species in the Mediterranean, with highest abundance in the Corso-ligurian basin and Gulf of Lyon; and sperm and Cuvier’s beaked whales are also regular species in the Mediterranean Sea, but less common than fin whale [71]. In contrast, the right whale specimen indicates a possible change in the regional whale composition. Indeed, not only this species is currently absent from the Mediterranean, it is also extremely rare in the historical record, with only three known records (Italy 1877, Alger 1888, and Sardinia 1991) [72]. In the archaeological record, there is an indirect proof of its prior presence at the entrance of Gibraltar: several plates of two barnacle species specific to right whales found in the Upper Magdalenian layers of cave in Málaga, Southern Spain [73]. The bone specimen identified in this study is thus the first direct archaeological evidence of right whale in the Mediterranean Sea. Given that the likelihood of vagrant individuals ending in the archaeological record is small, this result (combined with Málaga study) suggests that this species may have been regularly present in the Mediterranean before its near-extirpation from the eastern North Atlantic. Our results thus illustrate the importance of the zooarchaeological record for understanding the past distribution, abundance and ecology of whales.

4. Future perspectives: high-throughput methods

The need for accurate molecular identifications, coupled with the large proportion of unidentified archaeological and paleontological remains, emphasizes the importance of high-throughput methods in future cetacean barcoding projects. Traditional mtDNA barcoding approaches are well established, and typically provide robust species identifications (with the exception of cross-species hybrids [39]) for modern, degraded, and ancient samples. However, the need for careful sample preparation, clean-room extraction, and replicability when working with ancient remains can significantly increase the laboratory time and associated costs when working with many hundreds of remains. ZooMS, on the other hand, can more easily be scaled up for large datasets: using a plate approach, up to 96 samples can be processed at one time [74], potentially allowing for up to 1000 samples to be analysed per week [75]. Although ZooMS is a cost-effective, high-throughput screening method for large sample sets, it often lacks the taxonomic precision offered by genetic analysis. Given the importance of accurate molecular identifications for ancient whale bones and the large proportion of unidentified archaeological and paleontological remains, the future will likely rely on next-generation sequencing (NGS) approaches, which can offer both taxonomic precision and bulk processing. Here, we review the advantages and limitations of NGS methods, including hybridisation capture approaches, and their application to modern and ancient ecological studies.

(a) Next-Generation-Sequencing Methods

The advent of high-throughput or NGS methods has revolutionized the application of ancient genetics, massively enhancing the ability to recover ancient DNA templates from degraded remains. Although whole-genome ‘shotgun’ approaches have been attempted for species identification (most notably to refine the systematics of ancient hominids [76,77]), this approach is limited by the generally low percentage of endogenous DNA in ancient remains and the lack of nuclear reference genomes in public databases like GenBank or Ensembl [78]. Until comprehensive genome databases are available, mitochondrial genes and genomes and informative nuclear genes will primarily be the marker of choice for ancient cetacean identification, with DNA target enrichment followed by NGS as the most feasibly high-throughput method for data acquisition [79]. Enrichment (or capture) methodologies immobilize the target DNA regions through hybridisation to single-stranded DNA or RNA probes with high sequence homology (Figure 4). Following DNA extraction and library preparation, custom probes are used to immobilise the target DNA either on a solid-phase (e.g. surface of a microarray) or in-solution using biotinylated baits. Non-homologous DNA templates are then washed away, the target DNA is eluted off the probes and sequenced using NGS methods. Hybridisation capture of entire mitochondrial genomes (mitogenomes) has become increasingly common for ancient or degraded DNA studies [80–82], as it allows targeting of DNA regions up to several megabases, and can be scaled up for population level analyses.

Enrichment approaches are particularly useful for increasingly old samples, or those from tropical climates where preserved DNA templates may be degraded beyond the fragment length feasible for traditional PCR amplicons [83,84]. Moreover, complete mitogenomes have been shown to provide more robust topologies and estimates of divergence times than shorter mitochondrial sequences [85,86]. While NGS approaches are still considerably more expensive than capillary sequencing when dealing with small numbers of samples, they can be far more

cost-effective if designed in a way that minimizes unusable sequence (e.g. environmental contamination, non-target DNA), and captures information for the maximal number of samples [87]. Hybridization probes can be designed to capture and simultaneously sequence up to 100 specimens on a single lane of NGS instrumentation significantly reducing the per-sample costs, and providing mitogenome data for both initial species identification and subsequent phylogenetic analyses. Although the hybridization probes can be specifically designed to capture mitogenomes from single or multiple cetacean species, recent studies have demonstrated that ‘generic’ probes are capable of recovering mitogenomes from even phylogenetically distinct taxa [88–90]. Furthermore, palaeontological studies are already demonstrating the advantages of pairing ZooMS with NGS methods, screening thousands of bone fragments using ZooMS, followed by mitogenome capture of particular species of interest [75].

(b) Potential contributions of NGS to cetacean ecology

NGS molecular approaches are beginning to be applied more routinely to modern cetacean populations, recovering full mitogenomes [91–95], genomic single nucleotide polymorphisms (SNPs) [96,97] or even complete nuclear genomes [98] to develop more nuanced models of evolutionary systematics and population histories for various cetacean species. To date, hybridization capture has not yet been extensively applied to ancient marine species. The capture of ancient Steller’s sea cow nuclear genes [99], and ancient killer whale mitogenomes [100], however, demonstrate the utility of this approach for revealing both broad interordinal evolutionary systematics as well as more recent radiations. Molecular analyses targeting only fragments of mitochondrial DNA in paleontological and archaeological remains have already shed light on the past distribution and abundance of cetaceans over thousands of years, and the extent to which these populations have been impacted by humans [101]. Examples include recent studies on the gray whale (*E. robustus*) [102], bowhead whale (*Balaena mysticetus*) [103,104], North Atlantic right whale (*E. glacialis*) [46], and Hector’s dolphin (*Cephalorhynchus hectori*) [105] which have provided more accurate estimates of cetacean genetic diversity and population sizes prior to their overexploitation. These data provide a crucial baseline to conservation and management efforts, for example as part of IWC’s mandate to allow whale populations to recover to sustainable levels. Integrated with long-term climatic data and predictive habitat modelling, they can shed light onto how populations will respond to future anthropogenic change [68,70].

5. Conclusion

Molecular methods are already proving crucial to our understanding of the past distribution and abundance of whale species and much scope remains to expand their application to existing zooarchaeological collections. With further refinement of these methods and the augmentation of cetacean genomic reference datasets, we will be able to obtain increasingly fine-grained identifications to the subspecies, ecotype and population level. The systematic integration of well-dated archaeological and paleontological remains with high-throughput molecular analysis methods will reveal changes in habitat, genetic diversity, and population abundance associated with climatic and anthropogenic factors through millennial timescales [101,106].

Additional Information

Acknowledgments

We thank the following colleagues for supplying us with information and unpublished material: L. Pernet and M. Marco from the Musée site Lattara at Lattes (France); E. Gailledrat and P. Garmy, archaeologists at Lattara; C. Sanchez and G. Dupeyron archaeologist at Saint Martin Gruissan. We are also grateful for the input of the two anonymous reviewers, which significantly improved the quality of the manuscript.

Data Accessibility

DNA sequences: Genbank accessions KT923090-KT923101

Authors' Contributions

CS, KM, KR, LS carried out the molecular lab work, and participated in data analysis along with YvdH; MC, MH conceived of the study; CS, AC, AR, AG, BW participated in the design of the study and collected samples for analysis; CS and YvdH drafted the manuscript; all authors contributed to the manuscript and gave final approval for publication.

Competing Interests

We have no competing interests

Funding

This work was supported by European Research Council grant ORCA FP7-PEOPLE-2011-IIF 299075 to MH and MJC (PIs) and Labex Archimede programme IA- ANR-11-LABX-0032-01 at the University Paul-Valéry Montpellier 3 (AG). AR and AC were financially supported by the Project MORSE (CEP&S 2011-ANR-279 11-CEPL-006, French National Research Agency) and by the project “Y a-t-il eu une exploitation ancienne des baleines en Méditerranée?” (Maison des Sciences de l’Homme de Montpellier) along with AG.

References

1. Clark, G. 1947 Whales as an Economic Factor in Prehistoric Europe. *Antiquity* **21**, 84–104.
2. Monks, G., editor 2005 *The Exploitation and Cultural Importance of Sea Mammals*. Oxbow Books, Oxford.
3. Piana, E. 2005 Cetaceans and human beings at the uttermost part of America: a longlasting relationship in Tierra del Fuego. In *The exploitation and cultural importance of sea mammals* (ed G. Monks), pp. 121–137. Oxford: Oxbow Books.
4. Kandel, A. W. & Conard, N. J. 2003 Scavenging and processing of whale meat and blubber by Later Stone Age people of the Geelbek Dunes, Western Cape Province, South Africa. *South African Archaeological Bulletin* **58**, 91–93.
5. Mulville, J. 2002 The role of cetacea in prehistoric and historic Atlantic Scotland. *Int. J. Osteoarchaeol.* **12**, 34–48.
6. Dawson, P. C. 2001 Interpreting Variability in Thule Inuit Architecture: A Case Study from the Canadian High Arctic. *Am. Antiq.* **66**, 453–470.
7. Monks, G. G., McMillan, A. D. & Claire, D. E. S. 2001 Nuu-Chah-Nulth Whaling: Archaeological Insights into Antiquity, Species Preferences, and Cultural Importance. *Arctic Anthropol.* **38**, 60–81.
8. Pétilion, J.-M. 2008 First evidence of a whale-bone industry in the western European Upper Paleolithic: Magdalenian artifacts from Isturitz (Pyrénées-Atlantiques, France). *J. Hum. Evol.* **54**, 720–726.
9. Clapham, P. J., Young, S. B. & Brownell, R. L. 1999 Baleen whales: conservation issues and the status of the most endangered populations. *Mamm. Rev.* **29**, 37–62.
10. Scott Baker, C. & Clapham, P. J. 2004 Modelling the past and future of whales and whaling. *Trends Ecol. Evol.*

19, 365–371.

11. Reeves, R. R., Smith, T. D. & Josephson, E. A. 2007 Near-annihilation of a species: right whaling in the North Atlantic. In *The Urban Whale: North Atlantic Right Whales at the Crossroads* (eds S. D. Kraus & R. M. Rolland), pp. 39–74. Harvard University Press, Cambridge, Massachusetts, USA.
12. Lindquist, O., for Environmental History, C., Policy, of Stirling, U. & of St. Andrews, U. 2000 *The North Atlantic gray whale (Eschrichtius robustus): an historical outline based on Icelandic, Danish-Icelandic, English and Swedish sources dating from ca 1000 AD to 1792*. University of St. Andrews.
13. Alter, S. E., Rynes, E. & Palumbi, S. R. 2007 DNA evidence for historic population size and past ecosystem impacts of gray whales. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15162–15167.
14. Lyman, R. L. 2006 Paleozoology in the service of conservation biology. *Evol. Anthropol.* **15**, 11–19.
15. Smith, A. B. & Kinahan, J. 1984 The invisible whale. *World Archaeol.* **16**, 89–97.
16. Monks, G. G. 2005 An oil utility index for whale bones. In *The exploitation and cultural importance of sea mammals* (ed G. G. Monks), pp. 138–153. Oxbow Books, Oxford.
17. Savelle, J. M. 1997 The Role of Architectural Utility in the Formation of Zooarchaeological Whale Bone Assemblages. *J. Archaeol. Sci.* **24**, 869–885.
18. Szabo, V. E. 2008 *Monstrous fishes and the mead-dark sea: Whaling in the medieval North Atlantic*. Brill, Leiden and Boston.
19. Sabin, R. C. 2005 From the Palaeolithic to the present-day: the research value of marine mammal remains from archaeological contexts and the uses of contemporary museum reference collections. In *The Exploitation and Cultural Importance of Sea Mammals* (ed G. G. Monks), pp. 1–5. Oxbow Books, Oxford.
20. Buckley, M., Fraser, S., Herman, J., Melton, N. D., Mulville, J. & Pálisdóttir, A. H. 2014 Species identification of archaeological marine mammals using collagen fingerprinting. *J. Archaeol. Sci.* **41**, 631–641.
21. Huelsbeck, D. R. 1988 Whaling in the Precontact Economy of the Central Northwest Coast. *Arctic Anthropol.* **25**, 1–15.
22. van den Hurk, Y. 2014 A Whale of a Question: History and Social Implications of Cetacean Exploitation in the North Sea Area through Space and Time, MSc Thesis, University of Nottingham.
23. Dobney, K. M., Jaques, D., Barrett, J. & Johnstone, C. 2007 *Farmers, monks and aristocrats: the environmental archaeology of Anglo-Saxon Flixborough*. Oxbow Books.
24. Pétilion, J.-M. 2013 Circulation of whale-bone artifacts in the northern Pyrenees during the late Upper Paleolithic. *J. Hum. Evol.* **65**, 525–543.
25. Betts, M. W. 2007 The Mackenzie Inuit Whale Bone Industry: Raw Material, Tool Manufacture, Scheduling, and Trade. *Arctic* **60**, 129–144.
26. Higgs, N. D., Little, C. T. S. & Glover, A. G. 2011 Bones as biofuel: a review of whale bone composition with implications for deep-sea biology and palaeoanthropology. *Proc. Biol. Sci.* **278**, 9–17.
27. de Castilho, P. V. 2008 Utilization of cetaceans in shell mounds from the southern coast of Brazil. *Quat. Int.* **180**, 107–114.
28. Baker, C. S., Cipriano, F. & Palumbi, S. R. 1996 Molecular genetic identification of whale and dolphin products from commercial markets in Korea and Japan. *Mol. Ecol.* **5**, 671–685.
29. Baker, C. S. & Palumbi, S. R. 1994 Which whales are hunted? A molecular genetic approach to monitoring

whaling. *Science* **265**, 1538–1539.

30. Palumbi, S. R. & Cipriano, F. 1998 Species identification using genetic tools: the value of nuclear and mitochondrial gene sequences in whale conservation. *J. Hered.* **89**, 459–464.
31. Goldstein, P. Z. & DeSalle, R. 2011 Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *Bioessays* **33**, 135–147.
32. DeSalle, R., Egan, M. G. & Siddall, M. 2005 The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 1905–1916.
33. Dalebout, M. L., Baker, C. S., Mead, J. G., Cockcroft, V. G. & Yamada, T. K. 2004 A comprehensive and validated molecular taxonomy of beaked whales, family Ziphiidae. *J. Hered.* **95**, 459–473.
34. Valsecchi, E. & Amos, W. 1996 Microsatellite markers for the study of cetacean populations. *Mol. Ecol.* **5**, 151–156.
35. Palsbøll, P. J., Bérubé, M., Skaug, H. J. & Raymakers, C. 2006 DNA registers of legally obtained wildlife and derived products as means to identify illegal takes. *Conserv. Biol.* **20**, 1284–1293.
36. Baker, C. S., Dalebout, M. L., Lavery, S. & Ross, H. A. 2003 www.DNA-surveillance: applied molecular taxonomy for species conservation and discovery. *Trends Ecol. Evol.* **18**, 271–272.
37. Dalebout, M. L., Lento, G. M., Cipriano, F., Funahashi, N. & Baker, C. S. 2002 How many protected minke whales are sold in Japan and Korea? A census by microsatellite DNA profiling. *Anim. Conserv.* **5**, 143–152.
38. Baker, C. S., Cooke, J. G., Lavery, S., Dalebout, M. L., Ma, Y.-U., Funahashi, N., Carraher, C. & Brownell, R. L., Jr 2007 Estimating the number of whales entering trade using DNA profiling and capture-recapture analysis of market products. *Mol. Ecol.* **16**, 2617–2626.
39. Cipriano, F. & Palumbi, S. R. 1999 Genetic tracking of a protected whale. *Nature* **397**, 307–308.
40. Pichler, F. B., Dalebout, M. L. & Baker, C. S. 2001 Nondestructive DNA extraction from sperm whale teeth and scrimshaw. *Mol. Ecol. Notes* **1**, 106–109.
41. Pimper, L. E., Remis, M. I., Natalie, R., Goodall, P. & Baker, C. S. 2009 Teeth and Bones as Sources of DNA for Genetic Diversity and Sex Identification of Commerson's Dolphins (*Cephalorhynchus commersonii*) from Tierra del Fuego, Argentina. *Aquat. Mamm.* **35**, 330–333.
42. Evans, S. et al. In press. Using combined biomolecular methods to explore whale exploitation and social aggregation in Hunter-Gatherer-Fisher society in Tierra del Fuego. *Journal of Archaeological Science: Reports*
43. Yang, D. Y. & Speller, C. F. 2006 Co-amplification of cytochrome b and D-loop mtDNA fragments for the identification of degraded DNA samples. *Mol. Ecol. Notes* **6**, 605–608.
44. Losey, R. J. & Yang, D. Y. 2007 Opportunistic Whale Hunting on the Southern Northwest Coast: Ancient DNA, Artifact, and Ethnographic Evidence. *Am. Antiq.* **72**, 657–676.
45. Sinding, M.-H. S., Gilbert, M. T. P., Grønnow, B., Gulløv, H. C., Toft, P. A. & Foote, A. D. 2012 Minimally destructive DNA extraction from archaeological artefacts made from whale baleen. *J. Archaeol. Sci.* **39**, 3750–3753.
46. Rosenbaum, H. C., Egan, M. G., Clapham, P. J., Brownell, R. L., Malik, S., Brown, M. W., White, B. N., Walsh, P. & Desalle, R. 2000 Utility of North Atlantic Right Whale Museum Specimens for Assessing Changes in Genetic Diversity. *Conserv. Biol.* **14**, 1837–1842.
47. McLeod, B. A., Brown, M. W., Moore, M. J., Stevens, W., Barkham, S. H., Barkham, M. & White, B. N. 2008 Bowhead Whales, and Not Right Whales, Were the Primary Target of 16th-to 17th-Century Basque Whalers in

the Western North Atlantic. *Arctic* **61**, 61–75.

48. UBAAL (Unité de Bioarchéologie Animalde de Lille) 2014 Exploitation du Rorqual commun au début de Xie s. près de Dunkerque. Available at: <http://perso.univ-lille3.fr/~touteslati/spip.php?breve5&lang=fr>>. Accessed Aug18, 2014.
49. Webster, J. & Oxley, D. 2009 Protein Identification by Peptide Mass Fingerprinting using MALDI-TOF Mass Spectrometry. In *The Protein Protocols Handbook* (ed J. M. Walker), pp. 1117–1129. Humana Press.
50. Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J. C. 2009 Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **23**, 3843–3854.
51. Collins, M., Buckley, M., Grundy, H. H., Thomas-Oates, J., Wilson, J. & van Doorn, N. 2010 ZooMS: the collagen barcode and fingerprints. *SpectroscopyEurope* **22**, 6.
52. Buckley, M. 2015 Ancient collagen reveals evolutionary history of the endemic South American ‘ungulates’. *Proc. Biol. Sci.* **282**. (doi:10.1098/rspb.2014.2671)
53. Welker, F. et al. 2015 Ancient proteins resolve the evolutionary history of Darwin’s South American ungulates. *Nature* **522**, 81–84.
54. Kirby, D. P., Buckley, M., Promise, E., Trauger, S. A. & Holdcraft, T. R. 2013 Identification of collagen-based materials in cultural heritage. *Analyst* **138**, 4849–4858.
55. Buckley, M., Witcher Kansa, S., Howard, S., Campbell, S., Thomas-Oates, J. & Collins, M. 2010 Distinguishing between archaeological sheep and goat bones using a single collagen peptide. *J. Archaeol. Sci.* **37**, 13–20.
56. Welker, F., Soressi, M., Rendu, W., Hublin, J.-J. & Collins, M. 2015 Using ZooMS to identify fragmentary bone from the Late Middle/Early Upper Palaeolithic sequence of Les Cottés, France. *J. Archaeol. Sci.* **54**, 279–286.
57. van Doorn, N. L., Hollund, H. & Collins, M. J. 2011 A novel and non-destructive approach for ZooMS analysis: ammonium bicarbonate buffer extraction. *Archaeol. Anthropol. Sci.* **3**, 281–289.
58. von Holstein, I. C. C., Ashby, S. P., van Doorn, N. L., Sachs, S. M., Buckley, M., Meiri, M., Barnes, I., Brundle, A. & Collins, M. J. 2014 Searching for Scandinavians in pre-Viking Scotland: molecular fingerprinting of Early Medieval combs. *J. Archaeol. Sci.* **41**, 1–6.
59. Harvey, V. L., Egerton, V. M., Chamberlain, A. T., Manning, P. L. & Buckley, M. 2016 Collagen Fingerprinting: A New Screening Technique for Radiocarbon Dating Ancient Bone. *PLoS One* **11**, e0150650.
60. Tebbutt, S. J., Stewart, R. E. & Hill, D. F. 2000 Isolation and characterisation of DNA from whale bone. *J. R. Soc. N. Z.* **30**, 365–371.
61. Morin, P. A. & Mccarthy, M. 2007 Highly accurate SNP genotyping from historical and low-quality samples. *Mol. Ecol. Notes* **7**, 937–946.
62. Arndt, U. M. 2011 Ancient DNA analysis of northeast Pacific humpback whale (*Megaptera novaeangliae*). PhD Thesis, Simon Fraser University.
63. Collins, M. J. et al. 2002 The survival of organic matter in bone: a review. *Archaeometry* **44**, 383–394.
64. Ottoni, C., Koon, H. E. C., Collins, M. J., Penkman, K. E. H., Rickards, O. & Craig, O. E. 2009 Preservation of ancient DNA in thermally damaged archaeological bone. *Naturwissenschaften* **96**, 267–278.
65. Allentoft, M. E. et al. 2012 The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc.*

Biol. Sci. **279**, 4724–4733.

66. Hofreiter, M., Paijmans, J. L. A., Goodchild, H., Speller, C. F., Barlow, A., Fortes, G. G., Thomas, J. A., Ludwig, A. & Collins, M. J. 2015 The future of ancient DNA: Technical advances and conceptual shifts. *Bioessays* **37**, 284–293.
67. Macé, M. 2003 Did the Gray whale, *Eschrichtius robustus*, calve in the Mediterranean ? *Lattara* **16**, 153–164.
68. Alter, S. E. et al. 2015 Climate impacts on transocean dispersal and habitat in gray whales from the Pleistocene to 2100. *Mol. Ecol.* **24**, 1510–1522.
69. Rastogi, T., Brown, M. W., McLeod, B. A., Frasier, T. R., Grenier, R., Cumbaa, S. L., Nadarajah, J. & White, B. N. 2004 Genetic analysis of 16th-century whale bones prompts a revision of the impact of Basque whaling on right and bowhead whales in the western North Atlantic. *Can. J. Zool.* **82**, 1647–1654.
70. Foote, A. D. et al. 2013 Ancient DNA reveals that bowhead whale lineages survived Late Pleistocene climate change and habitat shifts. *Nat. Commun.* **4**, 1677.
71. Reeves, R. R. & Notarbartolo di Sciara, G. 2006 *The Status and Distribution of Cetaceans in the Black Sea and Mediterranean Sea*. IUCN Centre for Mediterranean Cooperation.
72. Rodrigues, A. S. L., Horwitz, L. K., Monsarrat, S. & Charpentier, A. 2016 Ancient whale exploitation in the Mediterranean: species matters. *Antiquity* **in press**.
73. Álvarez-Fernández, E. et al. 2014 Occurrence of whale barnacles in Nerja Cave (Málaga, southern Spain): Indirect evidence of whale consumption by humans in the Upper Magdalenian. *Quat. Int.* **337**, 163–169.
74. Korsow-Richter, K. K., Wilson, J., Jones, A. K. G., Buckley, M., van Doorn, N. & Collins, M. J. 2011 Fish 'n chips: ZooMS peptide mass fingerprinting in a 96 well plate format to identify fish bone fragments. *J. Archaeol. Sci.* **38**, 1502–1510.
75. Brown, S. et al. 2016 Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci. Rep.* **6**, 23559.
76. Green, R. E. et al. 2010 A draft sequence of the Neandertal genome. *Science* **328**, 710–722.
77. Reich, D. et al. 2010 Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060.
78. Blow, M. J., Zhang, T., Woyke, T., Speller, C. F., Krivoschapkin, A., Yang, D. Y., Derevianko, A. & Rubin, E. M. 2008 Identification of ancient remains through genomic sequencing. *Genome Res.* **18**, 1347–1353.
79. Jones, M. R. & Good, J. M. 2016 Targeted capture in evolutionary and ecological genomics. *Mol. Ecol.* **25**, 185–202.
80. Briggs, A. W. et al. 2009 Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318–321.
81. Thalmann, O. et al. 2013 Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* **342**, 871–874.
82. Avila-Arcos, M. C. et al. 2011 Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci. Rep.* **1**, 74.
83. Dabney, J. et al. 2013 Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758–15763.
84. Meyer, M. et al. 2014 A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**,

403–406.

85. Duchêne, S., Archer, F. I., Vilstrup, J., Caballero, S. & Morin, P. A. 2011 Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS One* **6**, e27138.
86. Mueller, R. L. 2006 Evolutionary Rates, Divergence Dates, and the Performance of Mitochondrial Genes in Bayesian Phylogenetic Analysis. *Syst. Biol.* **55**, 289–300.
87. Godden, G. T. et al. 2012 Making next-generation sequencing work for you: approaches and practical considerations for marker development and phylogenetics. *Plant Ecol. Divers.* **5**, 427–450.
88. Slon, V., Glocke, I., Barkai, R., Gopher, A., Hershkovitz, I. & Meyer, M. 2015 Mammalian mitochondrial capture, a tool for rapid screening of DNA preservation in faunal and undiagnostic remains, and its application to Middle Pleistocene specimens from Qesem Cave (Israel). *Quat. Int.* (doi:10.1016/j.quaint.2015.03.039)
89. Hancock-Hanser, B. L., Frey, A., Leslie, M. S., Dutton, P. H., Archer, F. I. & Morin, P. A. 2013 Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Mol. Ecol. Resour.* **13**, 254–268.
90. Mayer, C. et al. 2016 BaitFisher: A software package for multi-species target DNA enrichment probe design. *Mol. Biol. Evol.* (doi:10.1093/molbev/msw056)
91. Morin, P. A. et al. 2010 Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res.* **20**, 908–916.
92. Vilstrup, J. T. et al. 2011 Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *BMC Evol. Biol.* **11**, 65.
93. Arnason, U., Gullberg, A. & Janke, A. 2004 Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene* **333**, 27–34.
94. Xiong, Y., Brandley, M. C., Xu, S., Zhou, K. & Yang, G. 2009 Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. *BMC Evol. Biol.* **9**, 20.
95. Alexander, A., Steel, D., Slikas, B., Hoekzema, K., Carraher, C., Parks, M., Cronn, R. & Baker, C. S. 2013 Low diversity in the mitogenome of sperm whales revealed by next-generation sequencing. *Genome Biol. Evol.* **5**, 113–129.
96. Morin, P. A. et al. 2015 Geographic and temporal dynamics of a global radiation and diversification in the killer whale. *Mol. Ecol.* **24**, 3964–3979.
97. Fernández, R. et al. 2015 A genomewide catalogue of single nucleotide polymorphisms in white-beaked and Atlantic white-sided dolphins. *Mol. Ecol. Resour.* (doi:10.1111/1755-0998.12427)
98. Zhou, X. et al. 2013 Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nat. Commun.* **4**, 2708.
99. Springer, M. S. et al. 2015 Interordinal gene capture, the phylogenetic position of Steller’s sea cow based on molecular and morphological data, and the macroevolutionary history of Sirenia. *Mol. Phylogenet. Evol.* **91**, 178–193.
100. Foote, A. D., Newton, J., Ávila-Arcos, M. C., Kampmann, M.-L., Samaniego, J. A., Post, K., Rosing-Asvid, A., Sinding, M.-H. S. & Gilbert, M. T. P. 2013 Tracking niche variation over millennial timescales in sympatric killer whale lineages. *Proc. Biol. Sci.* **280**, 20131481.
101. Foote, A. D., Hofreiter, M. & Morin, P. A. 2012 Ancient DNA from marine mammals: studying long-lived species over ecological and evolutionary timescales. *Ann. Anat.* **194**, 112–120.

102. Alter, S. E., Newsome, S. D. & Palumbi, S. R. 2012 Pre-whaling genetic diversity and population ecology in eastern Pacific gray whales: insights from ancient DNA and stable isotopes. *PLoS One* **7**, e35039.
103. Borge, T., Bachmann, L., Bjørnstad, G. & Wiig, O. 2007 Genetic variation in Holocene bowhead whales from Svalbard. *Mol. Ecol.* **16**, 2223–2235.
104. Ho, S. Y. W., Saarma, U., Barnett, R., Haile, J. & Shapiro, B. 2008 The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS One* **3**, e1615.
105. Pichler, F. B. & Baker, C. S. 2000 Loss of genetic diversity in the endemic Hector's dolphin due to fisheries-related mortality. *Proc. Biol. Sci.* **267**, 97–102.
106. Hofman, C. A., Rick, T. C., Fleischer, R. C. & Maldonado, J. E. 2015 Conservation archaeogenomics: ancient DNA and biodiversity in the Anthropocene. *Trends Ecol. Evol.* **30**, 540–549.

Tables

mtDNA taxonomic identifications of Mediterranean archaeological cetacean bones. Samples listed in bold indicate those previously identified as gray whale remains through anatomical methods [67]; additional detail provided in Supplementary Table 2. The identified species are: right whale (*Eubalaena glacialis*), fin whale (*Balaenoptera physalus*), sperm whale (*Physeter catodon*), and Cuvier's beaked whale (*Ziphius cavirostris*).

Lab Code	Archaeological site	Chronology	DNA species ID	Collagen PMF (ZooMS) ID
WH501	Saint Martin, S. France	Late Antiquity	Right whale	Right/Bowhead whale
WH502	Cougourlude, S. France	Roman	no amplification	no ID
WH503	Cougourlude, S. France	Roman	Fin whale	no ID
WH504	Cougourlude, S. France	Roman	no amplification	no ID
WH505	Saint Sauveur, S. France	Roman	Fin whale	Fin whale
WH506	Saint Sauveur, S. France	Iron Age	Fin whale	Fin whale
WH507	Saint Sauveur, S. France	Iron Age	Fin whale	Fin whale
WH508	Saint Sauveur, S. France	Iron Age	Sperm whale	Sperm whale
WH509	Saint Sauveur, S. France	Iron Age	no amplification	Baleen whale (Mysticeti)
WH510	Saint Sauveur, S. France	Late Antiquity	Fin whale	Fin whale
WH511	Saint Sauveur, S. France	Iron Age	Fin whale	Fin whale
WH512	Saint Sauveur, S. France	Iron Age	Fin whale	Fin whale
WH513	Saint Sauveur, S. France	Iron Age	Fin whale	Fin whale
WH801	Nuraghe Lu Brandali, Sardinia	Bronze Age	Cuvier's beaked whale	Beaked Whale
WH802	Porto Torres, Sardinia	Roman	no amplification	Fin whale
WH803	Villa Sant'Imbenia, Sardinia	Early Middle Age	Fin whale	Fin whale
WH804	San Rocchino, Tuscany, Italy	Iron Age	no amplification	Fin whale

Figure captions

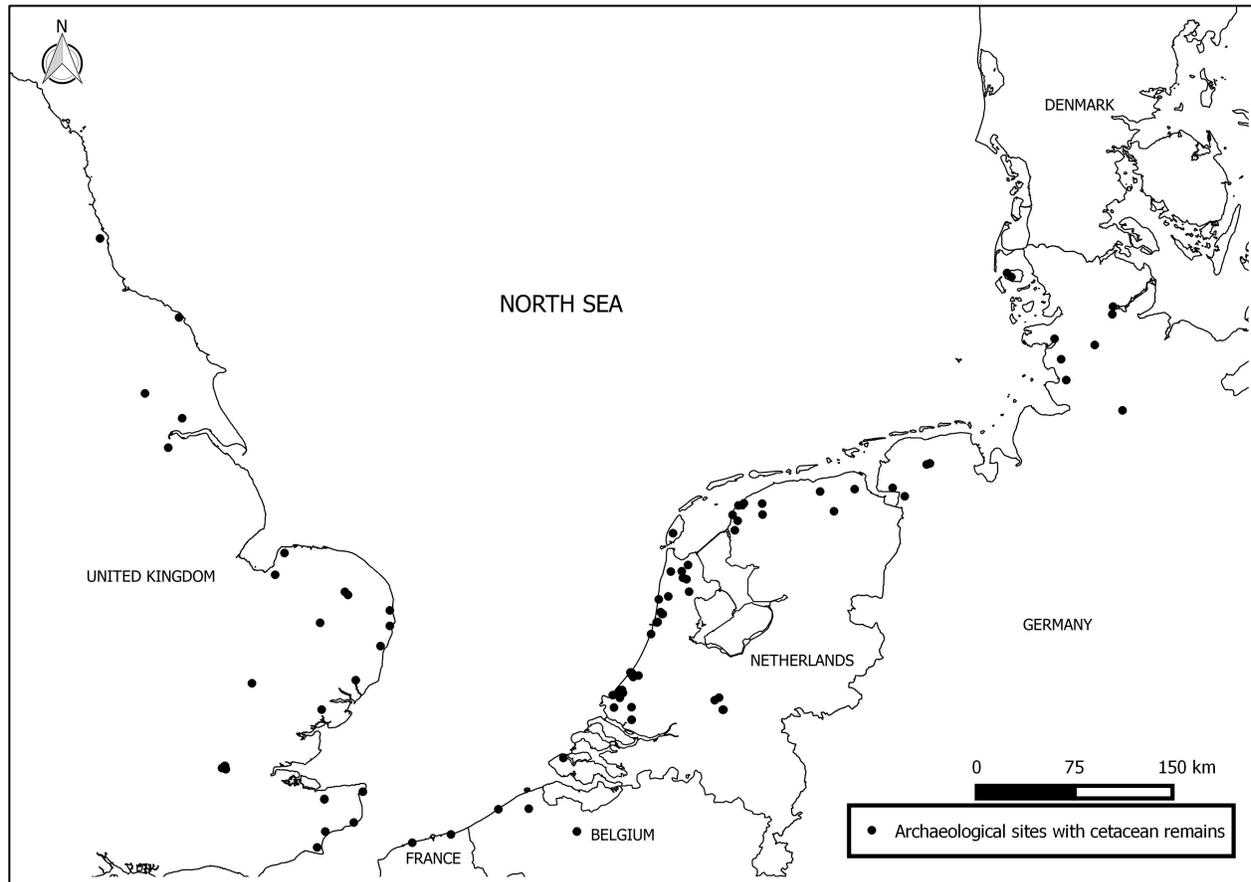


Figure 1. Location of southern North Sea archaeological sites with cetacean remains (including the east coast of England (n=27), the French region of Nord-Pas-de-Calais (n=2), Belgium (n=4), the Netherlands (n=56) and the North Sea coast of Germany (n=13).

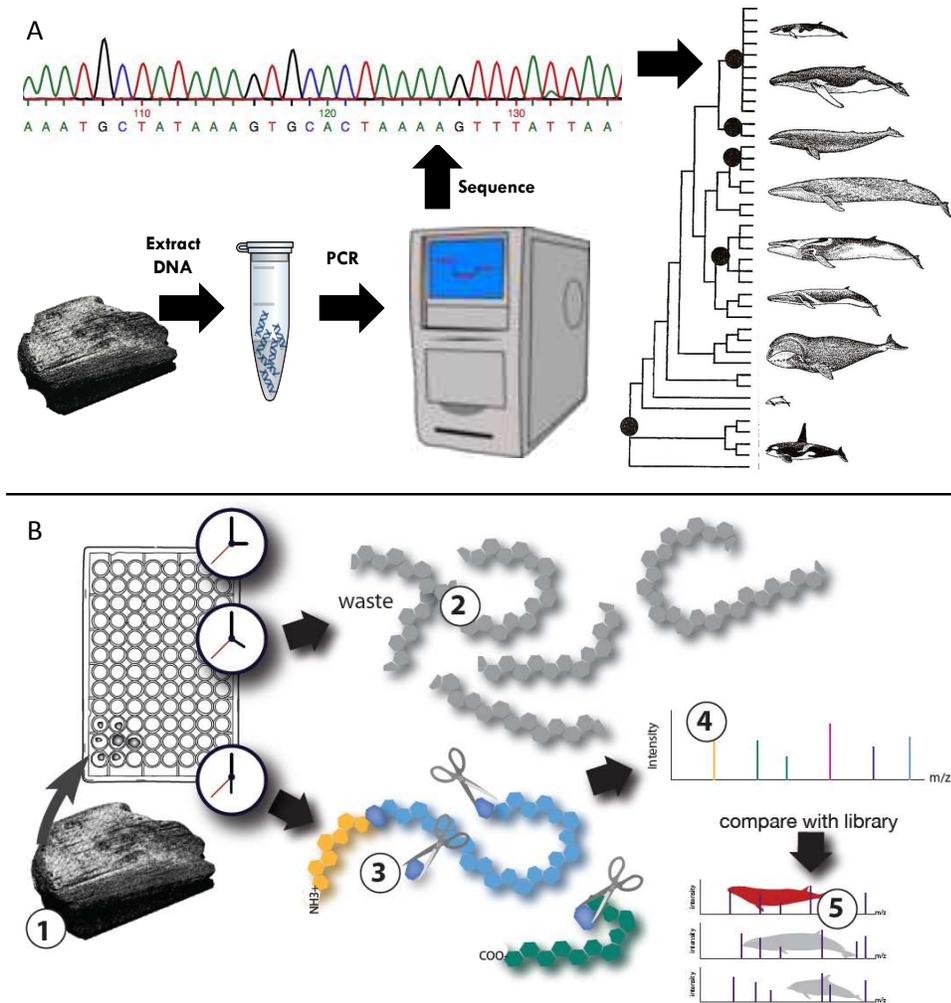


Figure 2: Two established methods for the molecular identification of ancient cetacean remains: A) DNA barcoding: DNA is extracted from the sample in a clean room, and PCR amplified targeting short fragments of mtDNA. Resulting sequences are compared a databank of known sequences for taxonomic identification. B) ZooMS: Samples are (1) demineralised in a weak acid solution; (2) collagen is gelatinised by heating at 65°C in an ammonium bicarbonate buffer, the collagen is then (3) enzymatically cleaved into peptides which are spotted with a matrix onto a target plate. The masses of the peptides are measured following desorption/ionisation of the sample using laser energy (MALDI) and (4) the peptide masses estimated by time of flight (TOF). The presence of specific peptides (5) is used for taxonomic identification.

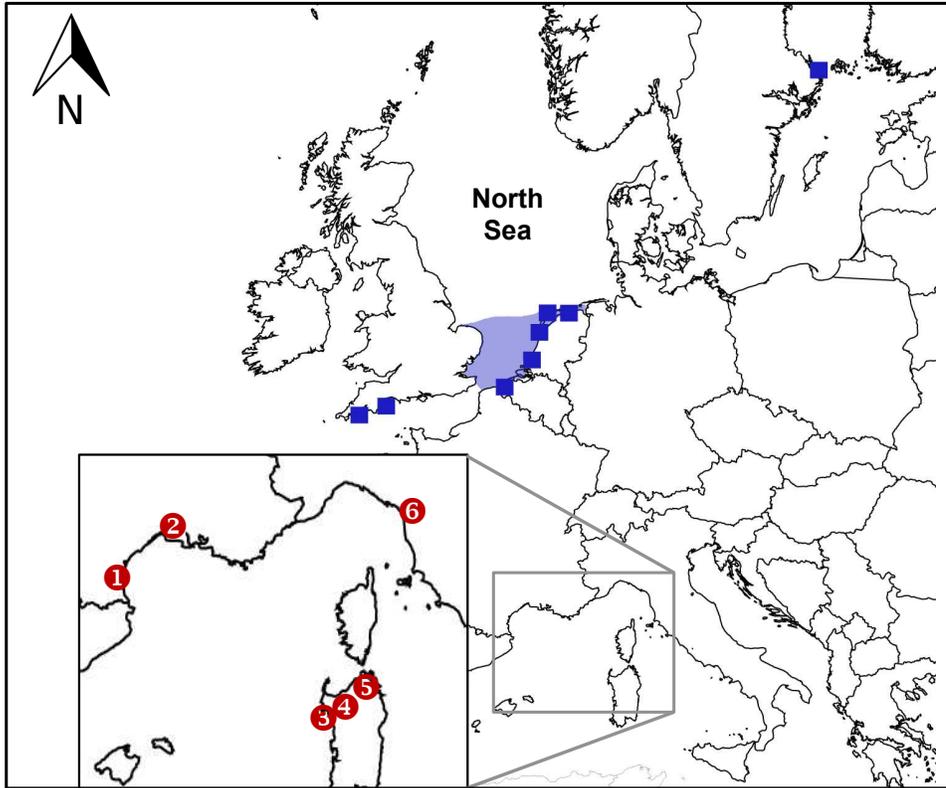


Figure 3: Map displaying locations of confirmed [68] paleontological gray whale finds in the northeast Atlantic (blue squares, and shaded area representing southern bight of the north sea) and the locations of the Mediterranean archaeological sites tested here (red circles:1) Saint Sauveur; 2) Cougourlude and Saint Martin; 3) Villa Sant'Imbenia; 4) Porto Torres; 5) Nuraghe Lu Brandali; 6) San Rocchino).

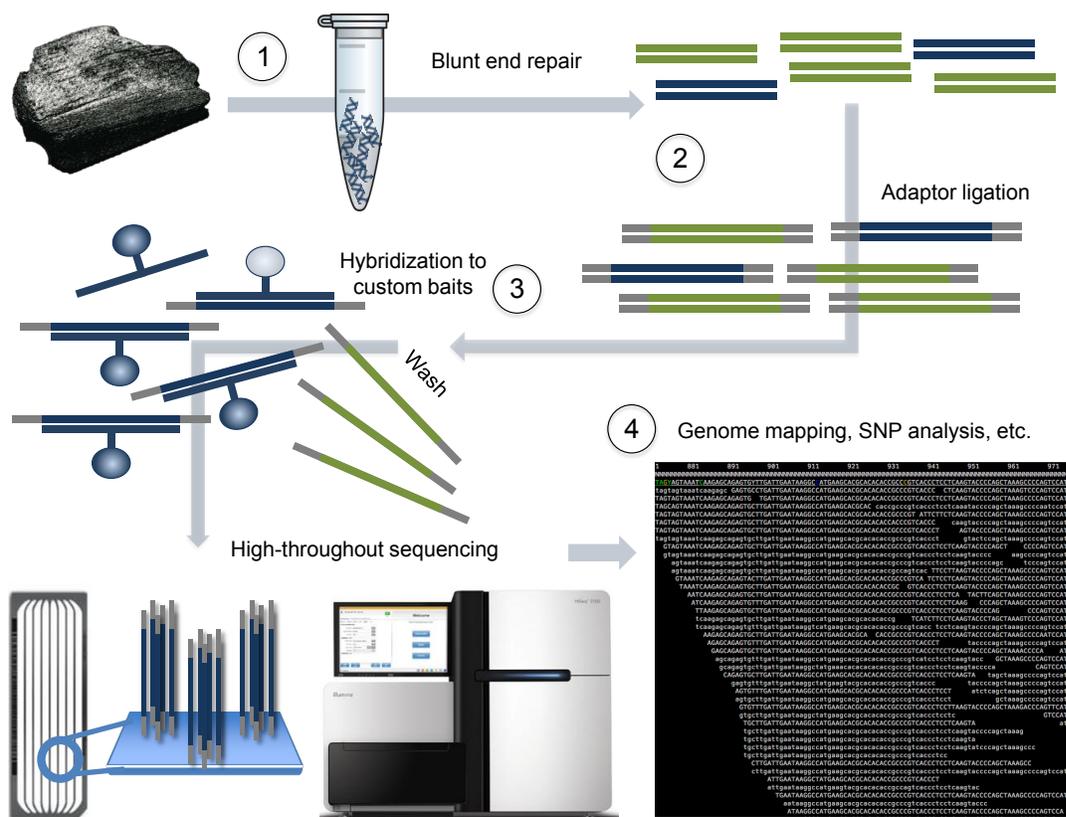


Figure 4: DNA hybridisation capture coupled to NGS: (1) DNA is extracted in a clean room and (2) NGS libraries are built from the extract; (3) the libraries are enriched for specific DNA sequences by hybridization to custom designed baits, and non-target templates are washed away; (4) enriched libraries are sequenced on an NGS platform, and (4) the resulting data are analyzed bioinformatically.

Supplementary material

Supplementary Material 1

Supplementary Table 1 Species identification summary of archaeological cetaceans remains from the North Sea Case study

Supplementary Table 2 Sample information, mtDNA and collagen PMF (ZooMS) results for the archaeological whale bones

Supplementary Table 1 Designated collagen peptide markers used for taxonomic identification of the archaeological cetacean samples

Supplementary Figure 1 Averaged MALDI-ToF mass spectra from sample WH509, which was identified only to the level of Mysticeti based on a lack of high-weight molecular markers necessary to differentiate fin whale, gray whale and humpback whale [20].

Supplementary Information

1. Materials and Methods

Biomolecular analysis was applied to 17 cetacean bones recovered from seven archaeological sites, including Saint Martin, Cougourlude, and Saint Sauveur on the southern coast of France [1,2], Nuraghe Lu Brandali, Porto Torres, Villa Sant'Imbenia, in Sardinia, and San Rocchino, Tuscany, Italy [3]. Based on previous morphological analysis, four of these samples from Saint Sauveur were presumed to represent possible gray whale remains [4] while the other could not be confidently assigned to species.

1.1 DNA sample preparation, extraction and amplification

The ancient whale samples were prepared and processed for DNA extraction in the Ancient DNA laboratory at University of York, following strict protocols for contamination control and detection, including positive pressure, the use of protective clothing, UV sources for workspace decontamination, and laminar flow hoods for extraction and PCR-set-up. Fragment of bone were immersed in 6% sodium hypochlorite for 5 mins, rinsed two times in HPLC grade water, UV irradiated for 30 min on two sides, and ground into powder. DNA from 20-55 mg of bone powder was extracted using a silica spin column protocol [5] as modified in Dabney et al. [6], and DNA was eluted in 50ul. PCR amplifications targeted a 182bp fragment of cytochrome b mitochondrial gene which has been demonstrated to successfully distinguish cetacean species [7,8]. PCR reactions and cycling conditions followed those described in Speller et al. [9]; successfully amplified products were sequenced using the forward primer at Eurofins Genomics, Ebersberg, Germany.

1.2 mtDNA sequence analysis and species identifications

ChromasPro software (www.technelysium.com.au) was used to visually analyse and edit the sequences and truncate primer sequences. Sequences were compared with published references through the GenBank BLAST application (<http://www.ncbi.nlm.nih.gov/BLAST/>), with multiple alignments of ancient and published reference sequences conducted using ClustalW [10], through BioEdit (<http://www.mbio.ncsu.edu/BioEdit>). Species identifications were assigned to a sample only if it was identical to published reference sequences from a single species in GenBank; species identities were further confirmed through 'DNA Surveillance', a web-based programme which provides robust cetacean identifications based on comparisons with a comprehensive set of validated cetacean reference sequences [11]. Twelve sequences were uploaded to the Genetic Sequence Database at the National Center for Biotechnical Information (NCBI) (GenBank ID:KT923090-KT923101).

1.3 Collagen peptide mass fingerprinting

The 17 cetacean samples were analyzed using the ZooMS protocol described in Buckley et al [12] and Evans et al. [8]. Between 10-30 mg of bone powder was fully demineralized through immersion in 0.6 M hydrochloric acid at room temperature or at 4°C. Samples WH505-507, WH511-513, and WH801-804 were centrifuged, the supernatant was discarded, and the samples rinsed three times with 200 µl AmBic solution (50 mMol ammonium bicarbonate, pH 8.0) before being gelatinised in 100 µl of AmBic solution for 1 hour at 65°C.

WH501-504 and WH508-510 underwent an additional ultrafiltration step. Following demineralization, these samples were centrifuged, the supernatant was discarded, and the collagen gelatinised through incubation in 250 µl of 0.6M HCl for three hours at 65°C. The collagen was ultrafiltered using Amicon Ultra-4 centrifugal filter units (30,000NMWL, EMD Millipore) until the supernatant was concentrated to approximately 100 µl. The retentate was washed three times with 200 µl AmBic solution, and concentrated to a final volume of 50 µl.

For all samples, the resulting collagen was incubated with 0.4µg of trypsin overnight at 37°C, acidified to 0.1% trifluoroacetic acid (TFA). The collagen was purified using a 100 µl C18 resin ZipTip® pipette

tip (EMD Millipore) with conditioning and eluting solutions composed of 50% acetonitrile and 0.1% TFA, while 0.1% TFA was used for the lower hydrophobicity buffer. The resulting collagen was eluted in 50 μ l.

1.4 Mass spectrometry and taxonomic identifications

One microlitre of the collagen extract was mixed with 1 μ l of α -cyano-hydroxycinnamic acid matrix solution (1% in conditioning solution) and spotted onto a 384 spot MALDI target plate, with calibration standards. Sample were spotted in triplicate, and run on a Bruker ultraflex III MALDI TOF/TOF mass spectrometer with a Nd:YAG smart beam laser. A SNAP averaging algorithm was used to obtain monoisotopic masses (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583), resulting in a total of 51 individual spectra.

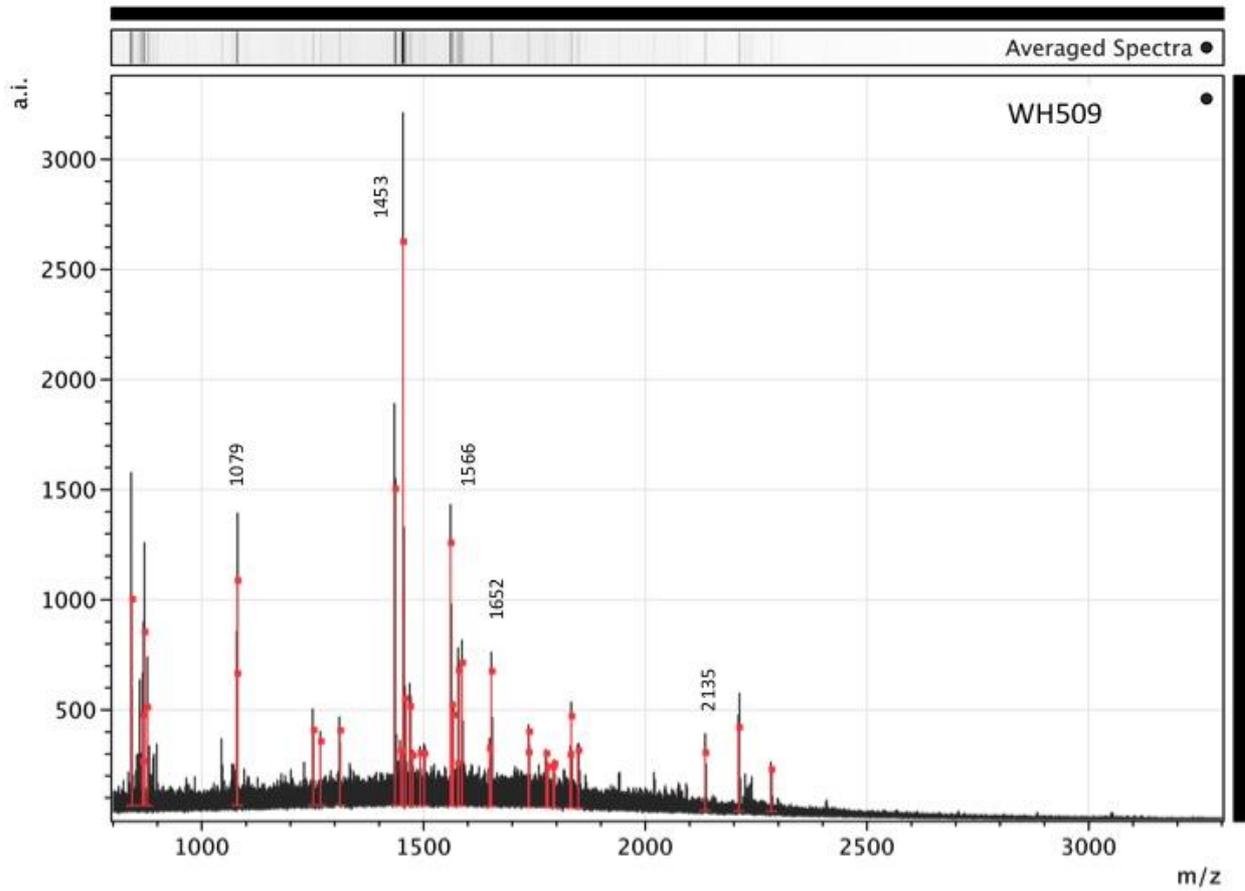
mMass software [13] was used to visually inspect the spectra; spectra from replicates of the same sample were averaged, and compared to the list of m/z markers for marine mammals presented in Buckley et al. [14] and Kirby et al. [15]. Taxonomic identifications were assigned at the most conservative level of identification (genus, or family level) based on the presence of unambiguous m/z markers.

2. Results

2.1 Taxonomic identifications

Following analysis of the mtDNA sequences and PMF spectra, taxonomic identifications could be assigned to 15 of the 17 samples. Taxonomic identifications were assigned to 12 archaeological samples using ancient mtDNA sequences and 14 samples using PMF spectra (Table S2; Table S3). The combined results produced 11 fin whale (*Balaenoptera physalus*), one sperm whale (*Physeter catodon*), one right whale (*Eubalaena glacialis*), one Cuvier's beaked whale (*Ziphius cavirostris*) and one family level identification (*Mysticeti*). ZooMS and mtDNA identifications were consistent for the 11 samples which produced results using both methods. The three samples that failed to amplify using the whale-specific cytb primers (WH502, 504, 509), also failed to produce unambiguous ZooMS identifications, suggesting poor overall biomolecular preservation in these samples.

Supplementary Figure 1: Averaged MALDI-ToF mass spectra from sample WH509, which was identified only to the level of Mysticeti based on a lack of high-weight molecular markers necessary to differentiate fin whale, gray whale and humpback whale ([Buckley et al. 2014](http://dx.doi.org/10.1111/j.1365-3113.2014.00581.x))



Supplementary Table 1: Species Identification summary of archaeological cetaceans remains from the North Sea Case study

Species	Skull Fragment	Maxilla	Mandible	Cervical Vertebra	Thoracic Vertebra	Lumbar Vertebra	Caudal Vertebra	Vertebra	Vertebral Epiphysis	Rib	Scapula	Humerus	Radius	Ulna	Phalanx	Partial Skeleton	Unidentified	Element not specified	TOTAL
Unknown Cetacean	1		3	1				14		1	1						201		221
Common Bottlenose Dolphin	2		1		12	8	9	2			2	1						136	173
Harbour Porpoise	12		1		3	7	13	16		1	2							32	87
Common Bottlenose Dolphin Cf.																		50	50
Large Cetacean								9		1									10
White Beaked Dolphin							1											9	10
Sperm Whale							1	3		2								3	9
Killer Whale					1	1	2		1			1	1	1				1	9
Unknown Baleen Whale	1		2							1							4		8
Northern Atlantic Right Whale	1						1	2								2			6
Unknown Dolphin	1							3									1		5
Unknown Odontoceti				3							1								4
Northern Bottlenose Whale						2	1												3
Common Minke Whale																		3	3
Common Minke Whale Cf.																		3	3
Unknown Beaked Whale								2											2
Long Finned Pilot Whale										1	1								2
Fin Whale										1								1	2
Blue/Fin Whale											2								2
Large Odontoceti																	1		1
Short Beaked Common Dolphin																		1	1
Common Bottlenose Dolphin/White Beaked Dolphin		1																	1
Long Finned Pilot Whale Cf.							1												1
Humpback Whale															1				1
Unknown Rorqual																	1		1
Sperm/Right/Humpback Whale						1													1
TOTAL	18	1	7	4	16	17	29	53	1	8	6	5	1	1	1	2	208	239	616

Supplementary Table 2: Sample information, mtDNA and collagen PMF (ZooMS) results for the archaeological whale bones

Lab Code	Archaeological site	Date	Evidence of processing	Morphological Identification	Bone	Fragment	Reference	DNA species ID	Collagen PMF (ZooMS) ID
WH501	Saint Martin, France	Late antiquity	carcass processing activity	Balaenidae	lumbar vertebra	processus lateral		<i>E. glacialis</i>	Right whale
WH502	Cougourlude, France	Roman	manufactured object		long bone	diaphysis		no amplification	Poor Spectra
WH503	Cougourlude, France	Roman	carcass processing activity	Balaenoptera	thoracic vertebra	processus dorsal		<i>B. physalus</i>	Poor Spectra
WH504	Cougourlude, France	Roman	carcass processing activity		lumbar vertebra	processus lateral		no amplification	Poor Spectra
WH505	Saint Sauveur, France	Roman	manufactured object	Eschrichtius or Balaenoptera	thoracic vertebra	body, C2 or C6	Macé 2003, fig 2	<i>B. physalus</i>	Fin Whale
WH506	Saint Sauveur, France	Iron Age II	no evidence			undetermined		<i>B. physalus</i>	Fin whale
WH507	Saint Sauveur, France	Iron Age II	carcass processing activity	Eschrichtius or Balaenoptera	vertebra	body	Macé 2003, fig. 3	<i>B. physalus</i>	Fin whale
WH508	Saint Sauveur, France	Iron Age II	carcass processing activity	Balaenidae or Eschrichtidae	rib	body	Macé 2003, fig. 5	<i>P. catodon</i>	Sperm Whale
WH509	Saint Sauveur, France	Iron Age II	no evidence	Eschrichtius or Megaptera	maxillar (upper)	blowhole fragment	Macé 2003, fig 4	no amplification	Mysticeti
WH510	Saint Sauveur, France	Late Antiquity/Early Middle Age	manufactured object	Eschrichtius	cervical vertebra	body	Macé 2003, fig 1	<i>B. physalus</i>	Fin whale
WH511	Saint Sauveur, France	Iron Age II	no evidence		vertebra	fragment		<i>B. physalus</i>	Fin whale
WH512	Saint Sauveur, France	Iron Age II	no evidence		unknown	undetermined		<i>B. physalus</i>	Fin whale
WH513	Saint Sauveur, France	Iron Age II	no evidence		unknown	undetermined		<i>B. physalus</i>	Fin whale
WH801	Nuraghe Lu Brandali, Sardinia	Bronze Age	Butchery marks		Vertebra	body		<i>Z. cavirostris</i>	Beaked Whale
WH802	Porto Torres, Sardinia	Roman	Butchery marks		unknown			no amplification	Fin whale
WH803	Villa Sant'Imbenia, Sardinia	Middle Ages (C7-8 AD)	Butchery marks		Scapula	?		<i>B. physalus</i>	Fin whale
WH804	San Rocchino, Italy	Iron Age	Butchery marks		Vetebra	body		no amplification	Fin whale

Supplementary Table 3: Designated collagen peptide markers used for taxonomic identification of the archaeological cetacean samples

Sample No.	(P1)	$\alpha 2(I)$ 988–1000 (A)	$\alpha 2(I)$ 494–508 (B)	$\alpha 2(I)$ 512–529 (C)	(P2)	$\alpha 2(I)$ 803–826 (D)	$\alpha 1(I)$ 602–634 (F)	$\alpha 2(I)$ 767–799 (G)	Collagen PMF ID
WH501	1079	1205?	1453	1566	1682	2135	2883	3023	Right whale
WH502	-	-	-	-	-	-	-	-	No identification
WH503	-	-	-	-	-	-	-	-	No identification
WH504	-	-	1453	-	1652	-	-	-	No identification
WH505	1079	1205?	1453	1566?	1652	2135	2883	3023	Fin Whale
WH506	1079	-	1453	-	1652	2135	2883	3023	Fin whale
WH507	1079	-	1453	1566?	1652	2135	2883	3023	Fin whale
WH508	1079	1205	1453	1550	1652	2133	2883	-	Sperm Whale
WH509	1079	-	1453	1566	1652	2135	2883?	-	Mysticeti
WH510	1,079	1,205	1,453	1,566	1,652	2,135	2,883	-	Fin whale
WH511	1079	1205?	1453	1566?	1652	2135	2883	3023	Fin whale
WH512	1079	1205	1453	1566?	1652	2135	2883	3023	Fin whale
WH513	1079	1205	1453	1566?	1652	2135	2883	3023	Fin whale
WH801	1063	-	1441	1550	-	2091?	-	-	Beaked whale
WH802	1079	1205	1453	1566?	1652	2135	2883	3023	Fin whale
WH803	1079	1205	1453	1566?	1652	2135	2883	3023	Fin whale
WH804	1079	1205	1453	1566?	1652	2135	2883	3023	Fin whale

? Indicates peak is present but at low intensity, or below signal to noise threshold

References:

1. Py, M. 2009 *Lattara, Lattes, Hérault. Comptoir gaulois méditerranéen entre Étrusques, Grecs, et Romains*. Éditions Errance.
2. Duperron, G. 2012 L'établissement antique de Saint-Martin-Le-Bas à Gruissan (Aude). In *Les ports antiques de Narbonne, vol. 2* (ed C. Sanchez), Programme collectif de recherche, ministère de la Culture, UMR 5140.
3. Wilkens, B. 2003 La Fauna sarda durante l'Olocene: le conoscenze attuali. (doi:10.1400/20528)
4. Macé, M. 2003 Did the Gray whale, *Eschrichtius robustus*, calve in the Mediterranean ? *Lattara* **16**, 153–164.
5. Yang, D. Y., Eng, B., Wayne, J. S., Dudar, J. C. & Saunders, S. R. 1998 Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* **105**, 539–543. (doi:10.1002/(SICI)1096-8644(199804)105:4<539::AID-AJPA10>3.0.CO;2-1)
6. Dabney, J. et al. 2013 Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758–15763. (doi:10.1073/pnas.1314445110)
7. Yang, D. Y. & Speller, C. F. 2006 Co-amplification of cytochrome b and D-loop mtDNA fragments for the identification of degraded DNA samples. *Mol. Ecol. Notes*
8. Evans, S. et al. In press. Using combined biomolecular methods to explore whale exploitation and social aggregation in Hunter-Gatherer-Fisher society in Tierra del Fuego. *Journal of Archaeological Science: Reports*
9. Speller, C. F., Hauser, L., Lepofsky, D., Moore, J., Rodrigues, A. T., Moss, M. L., McKechnie, I. & Yang, D. Y. 2012 High potential for using DNA from ancient herring bones to inform modern fisheries management and conservation. *PLoS One* **7**, e51122. (doi:10.1371/journal.pone.0051122)
10. Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680. (doi:10.1093/nar/22.22.4673)
11. Baker, C. S., Dalebout, M. L., Lavery, S. & Ross, H. A. 2003 www.DNA-surveillance: applied molecular taxonomy for species conservation and discovery. *Trends Ecol. Evol.* **18**, 271–272. (doi:10.1016/S0169-5347(03)00101-0)
12. Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J. C. 2009 Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **23**, 3843–3854. (doi:10.1002/rcm.4316)
13. Strohalm, M., Hassman, M., Kosata, B. & Kodíček, M. 2008 mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Commun. Mass Spectrom.* **22**, 905–908. (doi:10.1002/rcm.3444)
14. Buckley, M., Fraser, S., Herman, J., Melton, N., Mulville, J. & Pálisdóttir, A. 2014 Species identification of archaeological marine mammals using collagen fingerprinting. *J. Archaeol. Sci.* **41**, 631–641. (doi:10.1016/j.jas.2013.08.021)
15. Kirby, D. P., Buckley, M., Promise, E., Trauger, S. A. & Holdcraft, T. R. 2013 Identification of collagen-based materials in cultural heritage. *Analyst* **138**, 4849–4858. (doi:10.1039/c3an00925d)Text text text text text text.