Barcoding the largest animals on Earth: ongoing challenges and molecular solutions in the taxonomic identification of ancient cetaceans

Camilla Speller¹, Youri van den Hurk², Anne Charpentier³, Ana Rodrigues³, Armelle Gardeisen⁴, Barbara Wilkens⁵, Krista McGrath¹, Keri Rowse⁵, Luke Spindler¹, Matthew Collins¹ and Michael Hofreiter⁶

1BioArCh, Department of Archaeology, University of York, Environment Building, York, North Yorkshire YO10 5DD, UK
2Institute of Archaeology, University College London, 31–34 Gordon Square, London WC1H 0PY, UK
3CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE – CNRS, Montpellier Cedex 5, France
4Archéologie des Sociétés Méditerranéennes, UMR 5140, CNRS, Labex Archimede JA-ANR-11-LABX-0032-01, Université Paul-Valéry Montpellier, 34970 Lattes, France
5Dipartmento di Scienze della Natura e del Territorio, Università degli Studi, Sassari, Italy
6Institute of Biochemistry and Biology, Faculty of Mathematics and Natural Sciences, University of Potsdam, 14476 Potsdam, Germany

Over the last few centuries, many cetacean species have witnessed dramatic global declines due to industrial overharvesting and other anthropogenic influences, and thus are key targets for conservation. Whale bones recovered from archaeological and palaeontological contexts can provide essential baseline information on the past geographical 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 then discuss current opportunities presented by molecular identification methods such as DNA barcoding and collagen peptide mass fingerprinting (zooarchaeology by mass spectrometry), 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 hybridization capture followed by next-generation sequencing as cost-effective approaches for enhancing the ecological informativeness of these ancient sample sets.

This article is part of the themed issue ‘From DNA barcodes to biomes’.

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 nineteenth and early twentieth centuries), as well
as other anthropogenic influences (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 whales (Eubalaena glacialis) [11] and the Atlantic gray whale populations (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 an ecological priority for many countries. Developing informed conservation policies and sustainable management plans requires accurate historic data on cetacean abundance and distribution at various stages in their interactions with humans. Archaeological and palaeontological 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 maximize the information potential of the archaeological and palaeontological records.

In this article, we review the challenges with identifying ancient whale bones, as well as the opportunities provided by new molecular identification methods. We begin by summarizing 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

Despite the millennia of human–cetacean interactions, the research potentials of palaeontological and 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 with 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, its 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 more than 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, UK 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 eighteenth 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 (northwest 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, while 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 50 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 (electronic supplementary table S1).
Furthermore, among these taxonomically identified bones 119 originated from a single site, the Early Medieval site of Flixborough (represented by 115 common bottlenose dolphin Tursiops truncatus, three minke whale Balaenoptera acutorostrata and one 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 (less than 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 with smaller species (see discussion in [20]). Human behaviour may also preferentially increase fragmentation of large species compared with their smaller counterparts. Compared with dolphins or porpoises, baleen whales provide a more abundant supply of bone, with a 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. However, 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 (150–500 bp) of the control region or cytochrome b (cytb) gene to identify taxa and to estimate geographical 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 analysed 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 analysed 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 palaeontological or archaeological remains. As with modern whale product identification, studies of ancient specimens primarily concentrated on recovering short diagnostic fragments of mtDNA control region [40,41], cyt b 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 specialized laboratories (to prevent DNA contamination) remains a limiting factor to their large-scale application.

(b) Zooarchaeology by mass spectrometry: 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 analysed through matrix-assisted laser desorption/ionization 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]. As 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 analysed using a non-destructive ammonium bicarbonate buffer, which enables bone samples or artefacts to be analysed 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 bar-coding and ZooMS can offer robust taxonomic identifications of ancient cetaceans, 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 sub-tropical environments [65,66]. Despite 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; electronic supplementary material, table S2). 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 palaeontological 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 eighteenth century), are nearly all from the North Sea [68] (figure 3). The restricted spatial distribution of these bones is probably a poor reflection of their actual past range; indeed, habitat modelling predicts gray whales would have also occurred further south, including the Bay of Biscay, and to a lesser extent, the Mediterranean Sea [68]. Owing 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 (cyt b mtDNA analysis) and ZooMS (methods described in the electronic supplementary material), identifying 11 fin whale (*Balaenoptera physalus*), one sperm whale (*Physeter catodon*), one right whale (*E. glacialis*) and one Cuvier’s beaked whale (*Ziphius cavirostris*). The relative merits of both techniques in Table 1.

**Table 1.** 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 electronic supplementary material, table S2. The identified species are right whale (*Eubalaena glacialis*), fin whale (*Balaenoptera physalus*), sperm whale (*Physeter catodon*) and Cuvier’s beaked whale (*Ziphius cavirostris*).

<table>
<thead>
<tr>
<th>laboratory code</th>
<th>archaeological site</th>
<th>chronology</th>
<th>DNA species ID</th>
<th>collagen PMF (ZooMS) ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>WH501</td>
<td>Saint Martin, s. France</td>
<td>Late Antiquity</td>
<td>right whale</td>
<td>right/bowhead whale</td>
</tr>
<tr>
<td>WH502</td>
<td>Cougourlude, s. France</td>
<td>Roman</td>
<td>no amplification</td>
<td>no ID</td>
</tr>
<tr>
<td>WH503</td>
<td>Cougourlude, s. France</td>
<td>Roman</td>
<td>fin whale</td>
<td>no ID</td>
</tr>
<tr>
<td>WH504</td>
<td>Cougourlude, s. France</td>
<td>Roman</td>
<td>no amplification</td>
<td>no ID</td>
</tr>
<tr>
<td>WH505</td>
<td>Saint Sauveur, s. France</td>
<td>Roman</td>
<td>fin whale</td>
<td>fin whale</td>
</tr>
<tr>
<td>WH506</td>
<td>Saint Sauveur, s. France</td>
<td>Iron Age</td>
<td>fin whale</td>
<td>fin whale</td>
</tr>
<tr>
<td>WH507</td>
<td>Saint Sauveur, s. France</td>
<td>Iron Age</td>
<td>fin whale</td>
<td>fin whale</td>
</tr>
<tr>
<td>WH508</td>
<td>Saint Sauveur, s. France</td>
<td>Iron Age</td>
<td>sperm whale</td>
<td>sperm whale</td>
</tr>
<tr>
<td>WH509</td>
<td>Saint Sauveur, s. France</td>
<td>Iron Age</td>
<td>no amplification</td>
<td>baleen whale (Mysticeti)</td>
</tr>
<tr>
<td>WH510</td>
<td>Saint Sauveur, s. France</td>
<td>Late Antiquity</td>
<td>fin whale</td>
<td>fin whale</td>
</tr>
<tr>
<td>WH511</td>
<td>Saint Sauveur, s. France</td>
<td>Iron Age</td>
<td>fin whale</td>
<td>fin whale</td>
</tr>
<tr>
<td>WH512</td>
<td>Saint Sauveur, s. France</td>
<td>Iron Age</td>
<td>fin whale</td>
<td>fin whale</td>
</tr>
<tr>
<td>WH513</td>
<td>Saint Sauveur, s. France</td>
<td>Iron Age</td>
<td>fin whale</td>
<td>fin whale</td>
</tr>
<tr>
<td>WH801</td>
<td>Nuraghe Lu Brandali, Sardinia</td>
<td>Bronze Age</td>
<td>Cuvier’s beaked whale</td>
<td>beaked whale</td>
</tr>
<tr>
<td>WH802</td>
<td>Porto Torres, Sardinia</td>
<td>Roman</td>
<td>no amplification</td>
<td>fin whale</td>
</tr>
<tr>
<td>WH803</td>
<td>Villa Sant’Imbenia, Sardinia</td>
<td>Early Middle Age</td>
<td>fin whale</td>
<td>fin whale</td>
</tr>
<tr>
<td>WH804</td>
<td>San Rochino, Tuscany, Italy</td>
<td>Iron Age</td>
<td>no amplification</td>
<td>fin whale</td>
</tr>
</tbody>
</table>

![Figure 3.](image)

Figure 3. Map displaying locations of confirmed [68] palaeontological gray whale finds in the northeast Atlantic (filled squares, and shaded area representing southern bight of the North Sea) and the locations of the Mediterranean archaeological sites tested here (circled numbers: (1) Saint Sauveur; (2) Cougourlude and Saint Martin; (3) Villa Sant’Imbenia; (4) Porto Torres; (5) Nuraghe Lu Brandali and (6) San Rochino.)
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 (electronic supplementary material, 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'. Thus, none of the samples could be confirmed as gray whale using molecular approaches. These results illustrate the necessity of validating anatomical identifications using 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 seventeenth century Basque whaling remains determined that morphological identifications previously assigned to right whale were in fact bowhead whale [69]. Likewise, molecular analyses 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]—identifications 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 extant 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, although less common than fin whale [71]. By contrast, the right whale specimen indicates a possible change in the regional whale composition. Indeed, not only is this species currently absent from the Mediterranean, it is also extremely rare in the historical record, with only three known sightings (Italy 1877, Alger 1888 and Sardinia 1991) [72]. In the archaeological record, there is 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 a 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 up in the archaeological record is small, this result (combined with the 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 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 palaeontological 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 palaeontological remains, the future will probably 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 hybridization 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 markers of choice for ancient cetacean identification, with DNA target enrichment followed by NGS as the most feasible high-throughput method for data acquisition [79]. Enrichment (or capture) methodologies immobilize the target DNA regions through hybridization to single-stranded DNA or RNA probes with high sequence homology (figure 4). Following DNA extraction and library preparation, custom probes are used to immobilize 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. Hybridization capture of entire mitochondrial genomes (mitogenomes) has become increasingly common for ancient or degraded DNA studies [80–82], as it allows targeting of DNA even from highly degraded samples, 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 sequences (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 have already demonstrated the advantages of pairing ZooMS with NGS methods, by first screening thousands of bone fragments using ZooMS, followed by mitogenome capture of particular species of interest [75].

(b) Potential contributions of next-generation sequencing to cetacean ecology

NGS 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 mtDNA in palaeontological 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

Figure 4. DNA hybridization 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 a NGS platform, and the resulting data are analysed bioinformatically.
provide crucial baselines for conservation and management efforts, 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 levels. The systematic integration of well-dated archaeological and palaeontological 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].

References


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, Nottingham, UK.


26. Higgs ND, Little CTS, Glover AG. 2011 Bones as biofuel: a review of whale bone composition with


47. McLeod BA, Brown MWN, Moore MJ, Stevens W, Barkham SH, Barkham M, White BN. 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.


70. Foote AD et al. 2013 Ancient DNA reveals that bowhead whale lineages survived Late Pleistocene climate change and habitat shifts. Nat. Commun. 4, 1677. (doi:10.1038/ncomms2714)


