Quaternary Science Reviews 190 (2018) 137–147

Contents lists available at ScienceDirect

Quaternary Science Reviews

journal homepage: www.elsevier.com/locate/quascirev

Palaeoproteomics for human evolution studies

Frido Welker [a, b, *]

[a] Department of Human Evolution, Max-Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103, Leipzig, Germany
[b] Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350, Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 26 September 2017
Received in revised form
28 April 2018
Accepted 30 April 2018

Keywords:
Ancient proteins
ZooMS
Human evolution
Protein degradation
Hominin phylogenetics
Pleistocene

ABSTRACT

The commonplace sequencing of Neanderthal, Denisovan and ancient modern human DNA continues to revolutionize our understanding of hominin phylogeny and interaction(s). The challenge with older fossils is that the progressive fragmentation of DNA even under optimal conditions, a function of time and temperature, results in ever shorter fragments of DNA. This process continues until no DNA can be sequenced or reliably aligned. Ancient proteins ultimately suffer a similar fate, but are a potential alternative source of biomolecular sequence data to investigate hominin phylogeny given their slower rate of fragmentation. In addition, ancient proteins have been proposed to potentially provide insights into *in vivo* biological processes and can be used to provide additional ecological information through large scale ZooMS (Zooarchaeology by Mass Spectrometry) screening of unidentifiable bone fragments. However, as initially with ancient DNA, most ancient protein research has focused on Late Pleistocene or Holocene samples from Europe. In addition, only a limited number of studies on hominin remains have been published. Here, an updated review on ancient protein analysis in human evolutionary contexts is given, including the identification of specific knowledge gaps and existing analytical limits, as well as potential avenues to overcome these.

1. Introduction

The sequencing of Late Pleistocene hominin DNA has become commonplace 20 years after the first Neanderthal mtDNA sequences were published (Krings et al., 1997). Particularly the high-throughput sequencing revolution, the development of capture-based methods targeting ancient DNA, and tools to authenticate DNA sequences has made genomic research on extinct hominins possible on a large scale (Marciniak and Perry, 2017). Hallmarks in palaeoanthropology include the sequencing of multiple entire Neanderthal genomes (Green et al., 2010; Castellano et al., 2014; Prüfer et al., 2014; Hajdinjak et al., 2018) and the discovery of "Denisovans" as a sister-clade to Neanderthals (Krause et al., 2010; Meyer et al., 2012). Such datasets have laid the groundwork for detailed insights into past hominin population structure, size and movement (Kuhlwilm et al., 2016; Rogers et al., 2017), and the determination of nucleotide sequence variations unique to modern humans or Neanderthals, in turn providing access to a catalogue of phenotypically and physiologically relevant sequence mutations (Pääbo, 2014). The majority of these insights have been obtained through the analysis of hominin bones less than 0.1 million of years (Ma) old from temperate preservation conditions. No ancient hominin DNA older than 0.43 Ma has been recovered due to irreversible biomolecular degradation (Meyer et al., 2016).

Both theoretical models and experimental data, or a combination thereof, indicate that ancient DNA is highly fragmented by 1 Ma in optimal, cool, conditions, with exceedingly low concentrations of endogenous DNA content in temperate conditions already occurring after 0.1 Ma (Allentoft et al., 2012; Orlando et al., 2013; Hofreiter et al., 2015; Kistler et al., 2017). The process of DNA fragmentation should be more advanced at younger time scales in (sub)tropical conditions, leading to major biases in the spatiotemporal retrieval of ancient genomes (Slatkin and Racimo, 2016). As a result, and indicative of this fast-paced process, ancient DNA studies from hominin skeletal remains have largely reported the successful analysis of Late Pleistocene hominins (<120ka), with two pre-Eemian exceptions, Sima de los Huesos and Denisova Cave (Sawyer et al., 2015; Meyer et al., 2016; Slon et al., 2017). To these two a sample from Hohlenstein-Stadel can possibly be added as well (Posth et al., 2017). Inevitably, there will be future studies reporting ancient DNA from other Middle Pleistocene hominin specimens, but major evolutionary processes related to human

* Corresponding author. Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, Copenhagen, Denmark.
E-mail addresses: frido_welker@eva.mpg.de, frido.welker@snm.ku.dk.

https://doi.org/10.1016/j.quascirev.2018.04.033
0277-3791/

evolution in the (sub)tropics of Africa, the Middle East and Asia can currently be considered outside the realm of ancient DNA research. These include the formative processes around the appearance of *Homo sapiens* in Africa and the first incursions of this species into Eurasia (Hublin et al., 2017; Groucutt et al., 2018; Hershkovitz et al., 2018), the relationships among several Middle and Late Pleistocene hominin taxa in Eurasia, or the temporal and geographic spread of "Denisovans", in particular its hypothesized East Asian or Southeast Asian distribution (Cooper and Stringer, 2013; Browning et al., 2018).

Some researchers have therefore suggested focusing on ancient proteins instead, as these biomolecules degrade at a slower rate than DNA, can be recovered from the same tissues, and are phylogenetically informative among mammalian species (Welker et al., 2015a) and between hominin clades (Welker et al., 2016). Ancient proteins such as collagen type I (COL1) are also suitable to large-scale screening methods like collagen peptide mass fingerprinting (ZooMS; Buckley et al., 2009), providing access to novel hominin and fauna specimens of interest to ecological reconstructions, radiocarbon dating studies and genetic and proteomic analysis (Welker et al., 2015b; Harvey et al., 2016; Talamo et al., 2016; Charlton et al., 2016; Devièse et al., 2017). Ancient protein studies currently suffer from some of the same biases as ancient DNA studies, however. For example, most work published to date has focused on the Holocene and Late Pleistocene of Europe, with no shotgun proteomic study reported on Late Pleistocene (sub) tropical samples globally (Fig. 1). In addition to the gaps in our spatiotemporal understanding of protein survival, there remain significant lacunae in our knowledge of protein survival between different tissues, the way proteins could be used beyond phylogenetic applications, and differences in mass spectrometry measurements of protein degradation between MALDI-TOF MS and LC-MS/MS. This review aims to address these, at least in part, by providing a background to the current state of the field as well as future directions.

2. Methods in palaeoproteomics

Ancient protein survival has been studied since the 1950's (Abelson, 1954). Much of the work in the 20th century investigated mechanisms of degradation, in particular through amino acid racemization (AAR; Schroeder and Bada, 1976). This work forms the basis of our current understanding of protein degradation. However, attempts to directly sequence proteins using Edman degradation sequencing proved unsuccessful for ancient proteins, in particular because Edman sequencing requires highly purified, unmodified, and concentrated protein (Robbins et al., 1993). Such conditions are never met in ancient contexts. Alternatively, much work has gone into the immunological detection of ancient proteins but, like AAR and to a lesser extent MALDI-TOF MS, this does not allow direct observation of ancient protein sequences and/or sequence specific modifications (Brandt et al., 2002; Collins et al., 2003).

Currently, ancient proteins are therefore analyzed directly using two different mass spectrometry methods: MALDI-TOF MS and LC-MS/MS. Unlike previously used methods, both MALDI-TOF MS and LC-MS/MS ultimately rely on the presence and detection of SAPs (single amino acid polymorphisms) between homologous protein sequences of different genera, species or populations. These SAPs derive from nucleotide substitutions (SNPs, single nucleotide polymorphisms) at the genome level in the genetic sequence of protein-coding genes. The caused protein sequence variation allows the taxonomic or phylogenetic analysis of ancient proteins. This proteomic <> genomic link is unraveled in AAR or antibody-based studies. It therefore follows that MALDI-TOF MS and LC-MS/MS, particularly when employed in evolutionary frameworks, are tightly linked to evolutionary genomics and (bioinformatics) developments therein.

MALDI-TOF MS and LC-MS/MS differ in several significant aspects from each other (Fig. 2). For example, MALDI-TOF MS only provides insights on the total mass of individual peptides within an extract, while LC-MS/MS provides the additional benefit of obtaining the exact amino acid sequence of the present peptides. Accessible reviews including further details on the technical background to ancient shotgun and MALDI-TOF MS proteomic analysis have recently become available (Cappellini et al., 2014; Warinner et al., 2015; Dallongeville et al., 2016; Vinciguerra et al., 2016; Buckley, 2018), including the significance of understanding protein contamination risks (Hendy et al., 2018). The following therefore aims to reiterate some of the distinctive differences between MALDI-TOF MS and LC-MS/MS, and the way in which both methods have been applied to understand the Palaeolithic past.

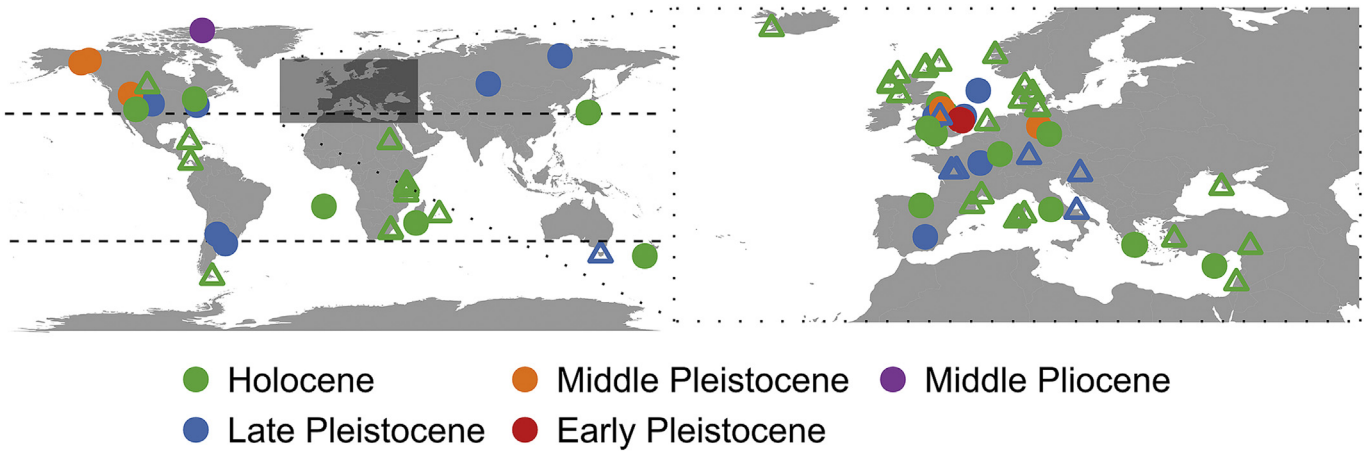

Fig. 1. The spatial distribution of published Pliocene, Pleistocene and Holocene ancient protein datasets from bone, dentine, and enamel. Circles represent shotgun proteome datasets and triangles COL1/ZooMS datasets. Preference was given to display proteomes when both COL1/ZooMS and proteome datasets are available from the same site. Horizontal lines indicate the boundaries of (sub)tropical zones between 35° north and south of the equator. Only studies where site locations are retraceable are included, and are roughly up-to-date as of March 2018.