ANNUAL REVIEWS

Ancient Biomolecules and Evolutionary Inference

Enrico Cappellini,[1,*] Ana Prohaska,[2,*]
Fernando Racimo,[1] Frido Welker,[3,4]
Mikkel Winther Pedersen,[2] Morten E. Allentoft,[1]
Peter de Barros Damgaard,[1] Petra Gutenbrunner,[5]
Julie Dunne,[6] Simon Hammann,[6,7]
Mélanie Roffet-Salque,[6] Melissa Ilardo,[1]
J. Víctor Moreno-Mayar,[1] Yucheng Wang,[1]
Martin Sikora,[1] Lasse Vinner,[1] Jürgen Cox,[5]
Richard P. Evershed,[6] and Eske Willerslev[1,2,8]

[1]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark; email: ecappellini@snm.ku.dk, ewillerslev@snm.ku.dk

[2]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, United Kingdom

[3]Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark

[4]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany

[5]Computational Systems Biochemistry, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

[6]Organic Geochemistry Unit, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom; email: r.p.evershed@bristol.ac.uk

[7]Department of Anthropology and Archaeology, University of Bristol, Bristol BS8 1UU, United Kingdom

[8]Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, United Kingdom

ANNUAL REVIEWS **Further**

Click here to view this article's online features:
• Download figures as PPT slides
• Navigate linked references
• Download citations
• Explore related articles
• Search keywords

*These authors equally contributed to the work.

Keywords

Abstract

Over the past three decades, studies of ancient biomolecules—particularly ancient DNA, proteins, and lipids—have revolutionized our understanding of evolutionary history. Though initially fraught with many challenges, today the field stands on firm foundations. Researchers now successfully retrieve nucleotide and amino acid sequences, as well as lipid signatures, from progressively older samples, originating from geographic areas and depositional environments that, until recently, were regarded as hostile to long-term preservation of biomolecules. Sampling frequencies and the spatial and temporal scope of studies have also increased markedly, and with them the size and quality of the data sets generated. This progress has been

made possible by continuous technical innovations in analytical methods, enhanced criteria for the selection of ancient samples, integrated experimental methods, and advanced computational approaches. Here, we discuss the history and current state of ancient biomolecule research, its applications to evolutionary inference, and future directions for this young and exciting field.

Contents

INTRODUCTION

Over the last few decades, studies of ancient biomolecules have transformed our understanding of the evolutionary history of life on Earth. Prior to this, evolutionary inferences had been drawn almost exclusively from molecular analyses of living organisms and the observation of phenotypic traits in fossils. However, such analyses provide only indirect evidence of the drivers and mechanisms that created present-day biodiversity. Ancient molecules, conversely, offer a direct window into the biological past and allow us to track evolutionary processes in real time. The categories of ancient molecules that have arguably made the biggest contribution to elucidating evolutionary history to date are nucleic acids, proteins, and lipids. While deoxyribonucleic acids (DNA) can dissect evolutionary processes with the highest resolution, proteins and lipids are important on longer temporal scales and in geographic areas that are less favorable to DNA preservation. In this

review, we introduce these three major categories of ancient biomolecules, summarize the history of their study, and discuss the foundations and frontiers of the field. We then go on to highlight their most important applications to evolutionary inference and outline where research is heading in the upcoming years.

HISTORICAL BACKGROUND

In 1984, DNA from a museum specimen of quagga, an equid species that went extinct in the nineteenth century, was successfully sequenced (1), thus marking the birth of the field of ancient DNA (aDNA) research. Since then, the focus of aDNA studies has progressed from small mitochondrial DNA fragments retrieved from a single species (2) to multiple species (3), to full genomic sequencing of one or a few specimens (4, 5), to single-nucleotide polymorphism capture-based population genomics (6) and whole-genome shotgun sequencing (7), often including over a hundred individuals. The age of specimens from which DNA can be successfully recovered has also increased significantly from a relatively modest couple of hundreds of years (1) to hundreds of thousands of years (8). In parallel to these developments in population genetics, studies of free extracellular DNA extracted directly from ancient depositional archives, such as sediments and ice cores (9, 10), known as environmental aDNA, have advanced from sequencing bacterial clones to metabarcoding, i.e., sequencing of amplicons using next-generation sequencing (NGS) platforms (11), to paleometagenomics, i.e., direct shotgun sequencing of ancient metagenomes from bulk samples. This has allowed the detection of a broad range of species from past environments at an unprecedented level of detail (12). Last, but not least, the field has also established a robust framework of operating procedures to effectively deal with contamination issues, which plagued earlier studies (reviewed in 13). These technological and methodological developments have allowed a rapid increase in scope and range of outputs over a relatively short period, generating fundamental new insights into the evolutionary past of a wide range of organisms across space and time.

The investigation of ancient protein residues, unlike aDNA, became possible only recently by application of a robust sequencing technique. This is rather surprising, given that the first successful attempt to characterize ancient protein residues dates back to 1954, when several amino acids were detected in fossil samples (14). In the following years, immunodetection methods were used to identify epitopes attributed to fossil proteins (15). However, the results produced in this pioneering phase were rarely conclusive, as the available methodologies were able to detect, but not sequence, ancient protein residues. The introduction of protein sequencing based on Edman degradation also had limited impact on the study of ancient proteins (16). Specimens suitable for this technique had to contain significant amounts of a single, purified, and minimally damaged protein—a set of conditions rarely met by ancient fossil remains. A turning point occurred in 2000, when mass spectrometry (MS) was applied to detect osteocalcin from ancient bone for the first time (17). The first significant application of this innovative approach to ancient specimens was the taxonomic identification of faunal remains based on collagen peptide mass fingerprinting (PMF) (18). The availability of instruments able to deliver progressively higher resolution and accuracy opened the way for reliable sequencing of ancient proteomes. In 2012, the first extended ancient bone proteome was retrieved from a ∼43,000-year-old mammoth (19). Since then, the approach has been successfully applied to a wide range of animal remains, timescales, and geographic regions to elucidate a number of evolutionary questions (20).

The study of ancient lipids in the modern era was very much stimulated by the emergence of gas chromatography (GC) and MS. The first analyses of lipids in geological settings were carried out in the 1960s, heralding the emergence of the field of organic geochemistry (21). Investigations of lipids in archaeological materials appeared in the 1970s (22). Over the following decades, studies

of lipids in ancient materials diversified into: (*a*) organic geochemistry, where lipids were used to reconstruct environments and processes in the geological past (23); (*b*) biomolecular paleontology, where the study of lipids was part of wider investigations of biomolecular preservation in fossils (24); and (*c*) biomolecular archaeology, where lipids were used to reconstruct human activities in the past (25). The past 50 years have witnessed an explosion in the application of ancient lipid analysis, ranging from fundamental studies of the evolution of life on Earth, petroleum exploration (26), paleoclimate reconstruction (26), biogeochemical cycles, and paleoecology (27), to studies of prehistoric human diet, agriculture, and many other cultural and social activities in the past (25). These diverse studies have provided the empirical and experimental framework for our current understanding of the range of sedimentary deposits, ecofacts, and artefacts that contain organic residues preserving lipid biomarkers.

FOUNDATIONS AND FRONTIERS

Sources of Ancient Biomolecules

Ancient DNA can be preserved in a variety of discrete remains, as well as a wide range of depositional archives (reviewed in 28). While bones and teeth remain the most widely used mineralized specimens for extracting aDNA, there is a wealth of other suitable calcified and mineralized substrates, such as eggshells, invertebrate shells, coprolites, and dental calculus, the latter two being particularly valuable for investigating ancient microbiomes (29). Keratinous material, e.g., hair, claws, and feathers, represents another important, though relatively scarce, source of animal aDNA (4, 30). Archaeobotanical remains, such as fossilized seeds, fruit, and wood, have been the dominant source of ancient plant DNA (reviewed in 31), while archaeological artefacts, e.g., lithics and ceramics, remain a promising, though highly contested and underutilized, source of aDNA originating primarily from food sources (28). Furthermore, environmental DNA, belonging to a diverse set of animal, plant, and bacterial species, can be extracted from various depositional contexts, including permafrost (9), cave and lake sediments (9, 11, 12, 32), and ice cores (10).

In addition to macrofossils and environmental deposits, there are a number of highly promising sources of aDNA that remain poorly explored owing to methodological and technological limitations. One notable example is microfossils, e.g., pollen, diatoms, and fungal spores, which represent a highly abundant fossil type in depositional environments across the globe. However, existing methods for the isolation of microfossils are mostly manual, e.g., selecting one microfossil at a time with micropipettes (33), and thus extremely time consuming.

Similar to aDNA studies, most ancient protein studies have focused on bones and teeth. The analysis of ancient bone proteomes has become a routine practice, with frequent identification of collagen type 1 (COL1), the most abundant component of the bone extracellular matrix. The stability and mechanical strength of collagen stem from its structure: a right-handed bundle of three tightly packed, parallel, left-handed polyproline II–type helices, and the presence of interstrand hydrogen bonds (34). Recently, dentine and enamel proteomes have also been investigated. The dentine and bone proteomes are similar in that they are both dominated by COL1 and their proteomes include hundreds of different proteins in present-day samples (35).

Enamel has its own advantages. It hosts a highly distinct proteome of ten or fewer proteins that are not found in dentine and can be analyzed relatively nondestructively. One of these proteins is amelogenin, which is expressed in humans in two isoforms from two different genes located on the nonrecombinant parts of the X and Y chromosomes (36). The confident identification of amelogenin Y-specific sequences provides a proteomics-based alternative to morphological and DNA-based sex determination techniques. Other sources, such as cultural heritage materials or

ancient dental calculus, have recently received considerable attention, as they provide additional information on human activity, health, and diet (29, 37). Ancient proteins are also abundant in mummified human and animal remains, as well as tissues derived from processed skin and hair, such as parchment (38), while investigation of ancient proteins from plant remains has, to date, been sporadic (39).

Like aDNA and ancient proteins, ancient lipids persist in a wide range of substrates, including ocean and lake sediments, sedimentary rocks, soils, fossil and subfossil remains, archaeological artefacts (e.g., pottery, stone tools), animal and human remains (soft tissue and bone lipids), coprolites, botanical remains, and a wide range of other deposits (tars, pitches, and bitumens) and hoards (or caches) of organic material such as bog butters (i.e., fatty substances, made either of animal carcass or dairy products found in peat bogs). A notable difference between lipids and other classes of biomolecules is that, under favorable preservation conditions, they are frequently recovered in high concentrations. For example, lipids comprise approximately 10% of the dry weight of peat, and the copious resin and bitumen deposits in Egyptian mummies are almost entirely lipid (40). Concentrations of absorbed lipids in archaeological vessels used for cooking are typically approximately 0.1 to 1 mg/g of the ceramic fabric (41), and bog butters have been recovered in multiple-kilogram quantities (42).
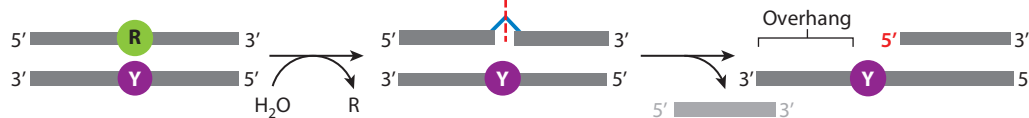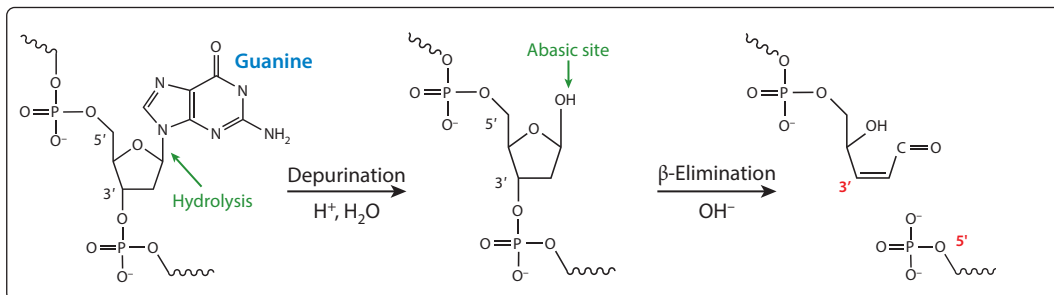
Characteristics of Ancient Biomolecules

The chemical structure of ancient biomolecules is heavily altered by a series of complex diagenetic reactions that begin upon the death of the organism and continue until their recovery. Detailed characterization of the typologies and the extent of chemical alterations occurring in ancient biomolecules is paramount for optimizing their recovery and authentication.
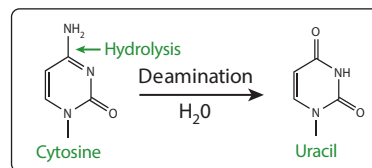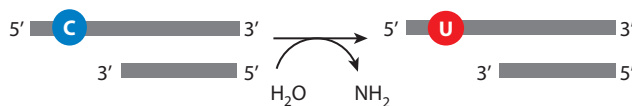
Ancient DNA. Ancient DNA is normally heavily fragmented and chemically modified. After the death of an organism, DNA is initially degraded by endogenous nucleases. This is soon followed by exogenous degradation processes, such as oxidation, hydrolysis, and background radiation, which alter the nitrogenous bases and cleave the sugar-phosphate backbone of the DNA molecules, leading to their destabilization and fragmentation (**Figure 1a**). There are four dominant types of aDNA damage: (*a*) fragmentation, (*b*) abasic sites (missing DNA bases), (*c*) cross-linking (condensation reactions between DNA and proteins or sugars), and (*d*) miscoding lesions (base pair modifications leading to the incorporation of incorrect bases during DNA amplification) (reviewed in 13). Fragmentation, abasic sites, and crosslinking all inhibit the amplification of aDNA, whereas miscoding lesions produce erroneous sequences that can significantly impact downstream analyses. Direct quantitative comparisons of aDNA fragmentation in a large number of bone samples from different geographic regions, time periods, and environments have revealed that the number of aDNA fragments exponentially decreases with the increase of their length, as the random breakage of long molecules results in an accumulation of shorter ones. While the rate of the fragmentation depends on different environmental factors, e.g., temperature, pH, and water availability, it appears to be initially rapid, most likely due to high enzymatic activity, and followed by reduced rates over the long term (43, 44). Moreover, high-throughput sequencing analyses of miscoding lesions have confirmed that (*a*) cytosine deamination to uracil, a thymine analog, is the most prominent base modification (45, 46); (*b*) this deamination increases toward fragment termini, where hydrolytic cleavage of phosphodiester bonds promotes the formation of single-stranded overhangs; and (*c*) depurination drives post mortem DNA fragmentation (13). Depurination preferentially occurs in adenines in younger samples, but in guanines in older samples, possibly reflecting differences in fragmentation dynamics between the two bases (46). With the introduction of
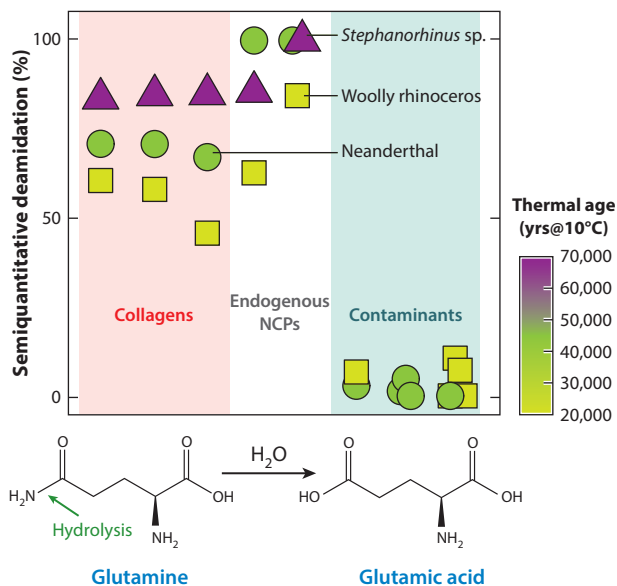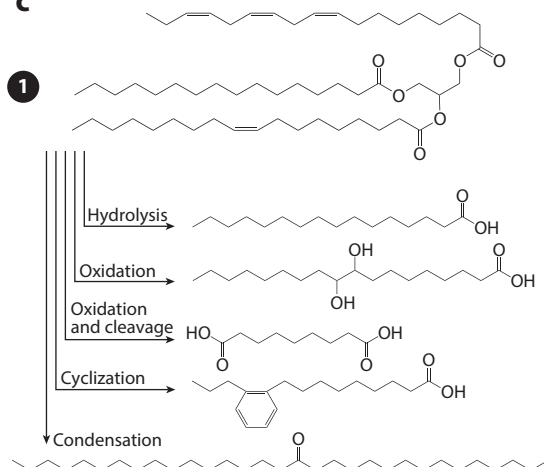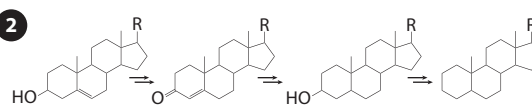
a

① Guanine → Depuration (H⁺, H₂O) → Abasic site → β-Elimination (OH⁻)

②

b

c

① Hydrolysis / Oxidation / Oxidation and cleavage / Cyclization / Condensation

②

(*Caption appears on following page*)

Figure 1 (*Figure appears on preceding page*)

Ancient biomolecules damage. (*a*) Ancient DNA (aDNA). ❶ Depurination and subsequent fragmentation, where the *N*-glycosyl bond between deoxyribose and a purine residue (adenine or guanine) is hydrolytically cleaved, leading to formation of an abasic site. This is often followed by the fragmentation of the DNA strand (single-strand breaks) through β-elimination, leaving 3′-aldehydic and 5′-phosphate ends (and 3′ and 5′ overhangs). ❷ Deamination of cytosine into uracil is the most common mechanism generating miscoding lesions in aDNA molecules, causing DNA polymerases to incorporate an adenine across from the uracil and resulting in cytosine-to-thymine and guanine-to-adenine substitutions. The chemical reactions and structures of the damage by-products are shown in boxes. Abbreviations: R, purine; Y, pyrimidine; C, cytosine; U, uracil. (*b*) Ancient proteins. Semiquantitative deamidation of common proteins identified in ancient bone proteomes. Values are based on spectral counting, including both glutamine and asparagine positions, in a Neanderthal (*circle*), woolly rhinoceros (*square*), and *Stephanorhinus* sp. (*triangle*) bone proteome. Chronological age is converted to thermal age to account for burial depth, latitude, and altitude, using a designated decision-support software tool (61). Primary data from References 47 and 48. Y-axis: 100% = full deamidation, 0% = no deamidation. Abbreviation: NCPs, noncollagenous proteins. (*c*) Ancient lipids. ❶ Ester lipids such as triacylglycerols are hydrolyzed, and liberated fatty acids can be oxidized, cleaved, or altered via cyclization or condensation mechanisms. In archaeological contexts, these reactions can also be human-induced; in particular, the formation of cyclic fatty acids and ketones requires excessive heating, e.g., during cooking (25). ❷ Lipids, such as sterols in sediments, undergo systematic alterations over millennia, subsequently losing double bonds and heteroatoms (R denotes an alkyl side chain). Similar degradation pathways exist for other lipids such as hopanoids or other terpenoids (49).

NGS, post mortem aDNA damage patterns became a key criterion for distinguishing endogenous sequences from contaminant DNA, i.e., DNA that penetrated the organism after its death and present-day DNA introduced through the excavation, storage, and handling of the ancient samples (13).

Particularly favorable environmental conditions for aDNA preservation include low temperatures, rapid desiccation, and high salt concentration. These factors facilitate destruction and/or inactivation of nucleases and reduce bacterial metabolic activity and hydrolytic attacks (50). Consequently, the oldest genomes sequenced to date come from specimens preserved under such conditions, for instance, a 110–130-kyr (thousands of years)-old bone of a polar bear in the Arctic Ocean Svalbard archipelago (51), a 700-kyr-old horse bone excavated in Yukon, Canada (8), and a 430-kyr-old hominin fossil found at Sima de los Huesos in Spain (52). Similarly, the oldest environmental aDNA has been sequenced from ice and permafrost (10) ranging between 400 and 800 kyr in age. In contrast, the age of the oldest environmental aDNA reads from the tropics is ~2 orders of magnitude lower (e.g., 53). While future studies may succeed in retrieving DNA sequences older than one million years, particularly from discrete specimens preserved under highly favorable conditions, current technological and methodological limitations make it hard to imagine such practice ever becoming routine.

Ancient proteins. A first evaluation of the decay of ancient proteins can be undertaken using sodium dodecyl sulfate (SDS) gel electrophoresis (19, 39). Unlike their modern counterparts, ancient protein extracts do not generate electrophoretic profiles with distinct, well-resolved bands characteristic of intact, freshly denatured proteins, but a continuous smear of protein residues. This is arguably due to spontaneous peptide backbone cleavage and the formation of covalent cross-links among proteins and possibly other organic compounds, such as carbohydrates and lipids. Protein fragmentation frequently occurs due to spontaneous nonenzymatic peptide backbone cleavage at the carboxyl side of asparagine (Asn) and glutamine (Gln) (54), and such fragmentation is correlated with their spontaneous and nonenzymatic deamidation to form aspartic acid (Asp) and glutamic acid (Glu), respectively (55). Peptide backbone cleavage also takes place at Asp and Glu residues, with rates higher than those for Asn and Gln (56).

Deamidation has been observed in almost all ancient protein studies undertaken to date, and it is generally higher in samples of greater thermal age (**Figure 1*b***) (57). Recent studies have observed that known contaminants present in ancient protein extracts exhibit very little to no

deamidation, while endogenous bone noncollagenous proteins (NCPs) generally have extensive deamidation, often close to 100%. In contrast, collagens display deamidation rates between those of common contaminants and those of NCPs. As a result, deamidation has been proposed as an effective method for distinguishing between endogenous and contaminating NCPs when extraction protocols permit it (47, 48).

Ancient protein residues also experience other forms of random, spontaneous, nonenzymatic alterations. Kynurenine, a product of tryptophan oxidation previously reported to spontaneously occur in situ in artificially aged wool, and aminoadipic acid, derived from lysine, are clear examples of age-dependent oxidative damage and were also observed in mammoth bone from the permafrost (19). Similarly, arginine and lysine carbonylation leads to missed cleavages during trypsin digestion, reducing peptide recovery in ancient samples (19). Carboxymethyllysine, an advanced glycation end product, was also observed in a moa bone sample not older than 1,000 years (58). Although recent work has focused on charting Gln and Asn deamidation, these and other important diagenetic pathways affecting protein preservation need to be examined in more detail.

As with aDNA, identifying the factors and the conditions favoring protein preservation over long temporal scales will guide the analysis of more challenging targets. Organo-metallic complexing, involving iron-catalyzed free-radical reactions facilitating protein cross-linking, has been put forward as one of the factors stabilizing soft tissue residues in fossil bone (59). Alternatively, stable covalent protein–protein bonds can also contribute to ancient bone protein preservation (60). Furthermore, it was recently demonstrated that it is the tight interaction between protein residues and the mineral matrix that plays a crucial role in ancient protein stabilization (61). Importantly, a growing body of experimental evidence and theoretical models indicate that ancient protein residues can be retrieved from considerably older epochs, even in geographic areas that are generally poorly suited for the preservation of ancient biomolecules (20, 61).

Ancient lipids. Lipids are the organic solvent–soluble components of living organisms, i.e., oils, fats, waxes, and resins. They are relatively recalcitrant because of the low abundances of functional groups and the dominance of saturated aliphatic chains, branches, and rings within their structure. This makes them inherently resistant to biodegradation and abiological decay compared to DNA and proteins, especially when they are protected within mineral or organic matrices. Entrapment in either organic or mineral matrices, e.g., sediment aggregates, pottery, bone, etc. (62), reduces the loss of biomolecules by diffusion and limits microbial activity by impeding access to lipid substrates. In highly dense lithified sedimentary materials, microbial activity is limited because of the reduced porosity and permeability restricting their access to water and essential nutrients. The degree of preservation of lipids is also highly dependent on the physicochemical conditions of the depositional environment, e.g., pH, redox potential, temperature (63), wetness (64), and biomass. The phenomenon known as sacrificial decomposition, which relies on the preferential decay of codeposited biological organic matter, serves to preserve lipids in various sedimentary deposits, artefacts, and ecofacts (65). The resistance of lipids to decay, combined with their persistence at the original site of deposition, makes them excellent candidates for use as biomarkers in molecular stratigraphic (i.e., chronological) investigations.

The preservation of lipids on geological timescales usually involves diagenetic defunctionalizations, e.g., hydrolysis of esters, reduction of double bonds, and loss of heteroatoms, most commonly oxygen (**Figure 1c**). The result is that lipid biomarkers preserved in very ancient geological sediments are generally hydrocarbon skeletons of the original biolipids. A classic paper by MacKenzie et al. (49) used the geological fate of sterols to illustrate the transformations of lipids from living organisms via a complex web of diagenetic and catagenetic reactions to yield geolipid assemblages. The transformations that occur are "systematic and sequential" (49, p. 491)

and the biochemical imprint in the structures, including stereochemistry, is evident even after hundreds of millions of years. Such observations apply to a wide range of the other classes of lipids, such as diterpenoids, triterpenoids, and porphyrins (23). A notable geological fate of lipids is their condensation with codeposited organic compounds to form geopolymers termed humic or kerogenous. Structurally recognizable lipid can be released from these materials by thermolysis and/or chemolysis.

More recent work on acyl lipids widely preserved at archaeological sites has provided a similar systematic understanding of the factors controlling deposition and degradation (66). Experimental studies of lipid behavior during food processing have provided insights into physicochemical processes involved in physical adsorption and thermally mediated transformations, as well as biodegradation during burial (66). A critical conclusion is that absorption of lipids into pores and onto clay surfaces of unglazed ceramic is vital to lipid survival, as this inhibits microbial degradation, explaining the survival of lipids in the oldest pottery investigated (67).
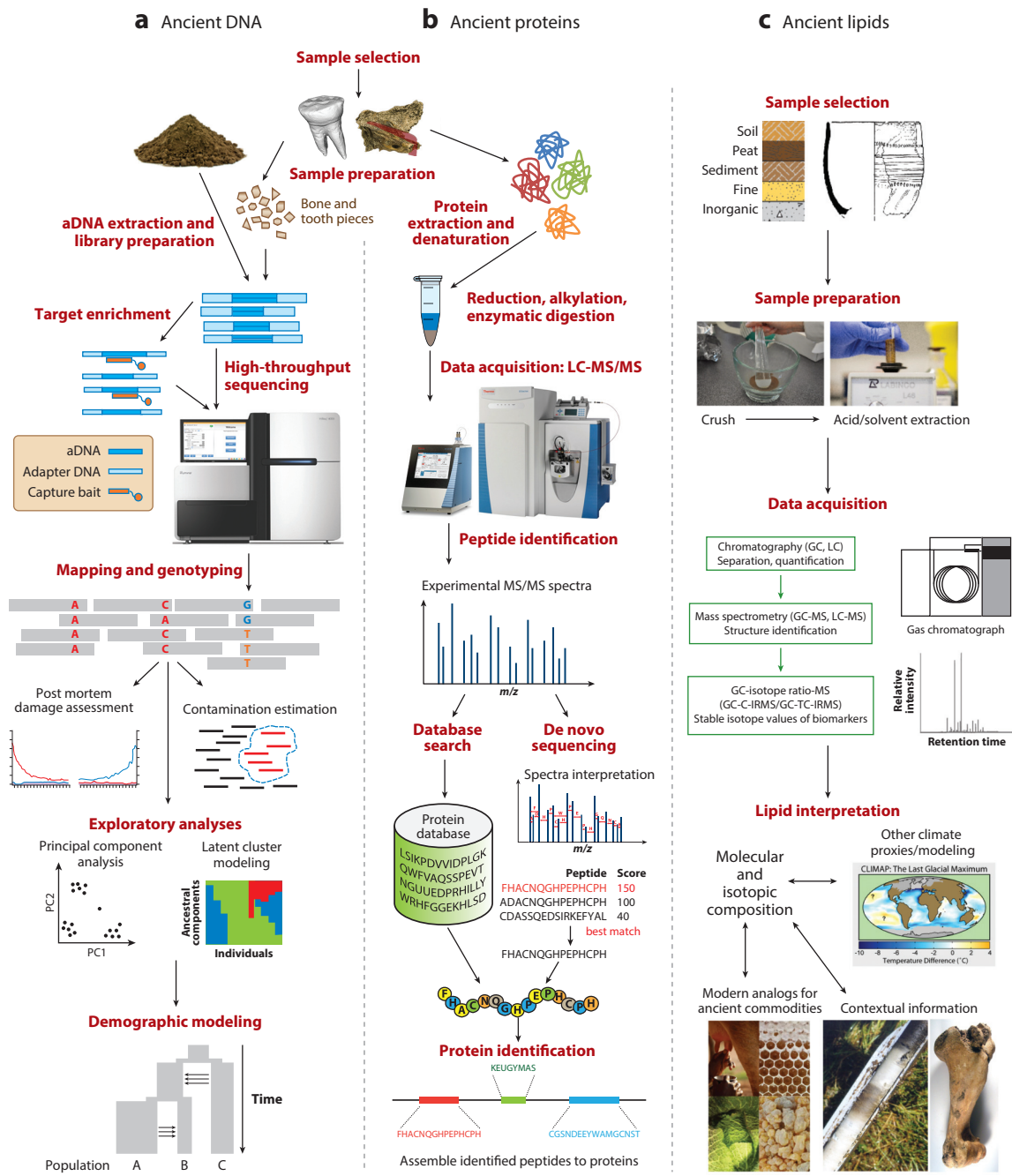
Analysis of Ancient Biomolecules

Every stage of ancient biomolecule investigation requires the adoption of dedicated procedures that take into account the key features of ancient biomolecules. Specimen selection, sample preparation, data generation, and interpretation all need to be adapted to deal with the limited amounts, the advanced degradation, and the extended contamination affecting the ancient biomolecules investigated.

aDNA. Over the last 30 years, a tremendous amount of effort has been invested into maximizing aDNA recovery. The introduction of high-throughput sequencing is arguably the single most important contributor to the expansion of the field. The associated explosion in data generation prompted a transformation of the procedures for their analysis and interpretation.

Sampling. The two key aims of aDNA retrieval (**Figure 2a**) are maximizing endogenous DNA content and minimizing contamination. Optimal sampling of ancient specimens is an important first step to achieve this. For instance, targeted sampling of the hardest and densest bone elements in the mammalian skeleton was instrumental in the emergence of ancient population genomics. The inner portion of the petrous part of the temporal bone in the skull, *pars petrosa*, and the cementum layer in tooth roots are considered to be the most favorable skeletal substrates for aDNA analysis owing to their high endogenous DNA contents (68, 69). Ethical, anthropological, and conservation concerns that have been recently raised against extensive sampling of petrous bones have prompted the development of less invasive approaches to retrieve bone powder from compact layers of such bones (70). For subsampling environmental aDNA, on the other hand, it is paramount to minimize the risk of contamination during sediment collection in the field and subsequent storage. For instance, spiking the surface of samples with detectable tracers helps with testing the penetration depth of contamination (9, 12).

Extraction. The characteristics of each sample substrate pose specific requirements for the extraction procedures, related to the effective digestion of the material and the solubilization of the DNA. For mineralized samples (e.g., bones, teeth), this is often undertaken with the aid of buffers; demineralizing agents, such as ethylenediaminetetraacetic acid (EDTA); detergents, such as SDS; surfactants, such as *N*-lauroyl-sarcosine; and proteinases. Collectively, these compounds lyse cell walls, degrade proteins, and release the DNA into solution. In case of sediment, lysis solution also breaks down the remaining organic structures and releases DNA molecules initially bound to the surface of mineral particles (71). The choice of lysis buffer depends on the type of sediment,

e.g., clay-rich (32) versus more organic (12, 72). With mineralized samples, the ratio of dissolved exogenous to endogenous DNA in a digestion buffer starts high and decreases with the time of digestion (73). Thus, an optimized predigestion step can efficiently remove the contaminant DNA without affecting the endogenous DNA. Following digestion, dissolved DNA is purified either by

a Ancient DNA

Sample selection

Sample preparation

Bone and tooth pieces

aDNA extraction and library preparation

Target enrichment

High-throughput sequencing

aDNA
Adapter DNA
Capture bait

Mapping and genotyping

Post mortem damage assessment

Contamination estimation

Exploratory analyses

Principal component analysis

Latent cluster modeling

PC2 / PC1

Ancestral components

Individuals

Demographic modeling

Time

Population A B C

b Ancient proteins

Sample selection

Protein extraction and denaturation

Reduction, alkylation, enzymatic digestion

Data acquisition: LC-MS/MS

Peptide identification

Experimental MS/MS spectra

m/z

Database search

De novo sequencing

Spectra interpretation

m/z

Protein database
LSIKPDVVIDPLGK
QWFVAQSSPEVT
NGUUEDPRHILLY
WRHFGGEKHLSD

Peptide	Score
FHACNQGHPEPHCPH	150
ADACNQGHPEPHCPH	100
CDASSQEDSIRKEFYAL	40

best match

FHACNQGHPEPHCPH

Protein identification

KEUGYMAS

FHACNQGHPEPHCPH   CGSNDEEYWAMGCNST

Assemble identified peptides to proteins

c Ancient lipids

Sample selection

Soil
Peat
Sediment
Fine
Inorganic

Sample preparation

Crush   Acid/solvent extraction

Data acquisition

Chromatography (GC, LC)
Separation, quantification

Mass spectrometry (GC-MS, LC-MS)
Structure identification

GC-isotope ratio-MS
(GC-C-IRMS/GC-TC-IRMS)
Stable isotope values of biomarkers

Gas chromatograph

Relative intensity

Retention time

Lipid interpretation

Molecular and isotopic composition

Other climate proxies/modeling

CLIMAP: The Last Glacial Maximum

Temperature Difference (°C)

Modern analogs for ancient commodities

Contextual information

(*Caption appears on following page*)

Figure 2 (*Figure appears on preceding page*)

Workflows for the analysis of ancient biomolecules. (*a*) Ancient DNA (aDNA). Sample material is carefully selected and prepared to maximize the proportion of endogenous DNA. Sequencing libraries are prepared from the extracted aDNA through the ligation of platform-specific adapters. High-throughput sequencing can be performed directly on library DNA or after enrichment (capture) using target-specific baits. Initial authentication of aDNA sequencing data includes assessment of post mortem DNA damage, analysis of fragment lengths, and estimation of contamination. Subsequently, the sequencing data can be used for exploratory analyses and modeling. (*b*) Ancient proteins. The analysis starts with demineralization of the bone/tooth mineral matrix of the selected samples. Protein residues are digested, usually with trypsin, after which they are purified and injected into an LC-MS/MS instrument. Experimental tandem MS (MS/MS) spectra are subsequently matched against established protein sequence databases, or analyzed de novo, in order to correctly infer peptide sequences and protein content of the ancient sample. (*c*) Ancient lipids. The selected samples are crushed into a powder and extracted using organic solvents. Chromatographic, mass spectrometric, and isotopic techniques are used to separate, identify, and characterize the compounds in these extracts. Results of the analyses are then interpreted within a multidisciplinary framework. Abbreviations: GC, gas chromatography; GC-C-IRMS, gas chromatography–combustion–isotope ratio mass spectrometry; LC, liquid chromatography; MS, mass spectrometry; TC, thermal conversion. Photo credits: Illumina Inc.; Thermo Fisher Scientific Inc.; Pinhasi et al. (177); Ken Chatterton, Water, Engineering, and Development Centre, Loughborough University; Wessex Archaeology; Freepik; and Wikimedia Commons.

precipitation [e.g., in phenol-chloroform extractions (74)] or through reversible adsorption [e.g., on silica, either in solution (8) or as a membrane in spin-columns]. The binding affinity of DNA molecules to silica is size dependent and can be further adjusted by changing salt concentration and pH (7). Sediments may require additional purification treatments to remove high levels of interfering organic compounds (e.g., 12). The use of multiple purification steps, however, represents a trade-off between removing enough inhibiting substances to allow downstream analysis and maintaining workable quantities of DNA.

Next-generation sequencing. NGS has now replaced Sanger sequencing in the aDNA field. By rendering the previously required transfer of plasmid libraries and bacterial cloning unnecessary, NGS tremendously increases the amount of retrievable data. The NGS workflow can be summarized as follows: first, building DNA sequencing libraries using DNA ligation technologies; second, amplifying libraries using polymerase chain reaction (PCR); third, performing massively parallel sequencing; and fourth, conducting downstream bioinformatics analyses. While multiple sequencing platforms have been used during the maturation period of ancient genomics, the Illumina platform has largely outcompeted the various other commercial options, primarily owing to its massive output of short DNA reads (13).

The two most widely used sequencing approaches today are shotgun and target-enriched sequencing. With shotgun sequencing, the extracted DNA is directly converted into a sequencing library. In contrast, target-enrichment by hybridization, a procedure commonly known as capture, selects for DNA library fragments of interest with either DNA or RNA baits. Capture may be mediated by baits on solid-phase microarrays (75, 76) or by biotinylated baits in solution (77). Each of the two sequencing approaches has its own advantages and limitations (**Table 1**). One of the main advantages of shotgun sequencing is that it ensures a complete and unbiased sequencing of the extracted DNA. However, owing to poor aDNA preservation and extensive microbial contamination, sufficient coverage of the endogenous genomic material requires a relatively high volume of shotgun sequencing, which can often be prohibitively expensive. Capture-enriched sequencing may vastly increase the proportion of on-target DNA reads (78–80), thereby considerably reducing sequencing costs (80, 81). Its main limitation is that only studies with overlapping target regions can be comprehensively compared. Additional challenges include effective hybridization of targeted sequences, especially with heavily degraded molecules (e.g., 80); reduced complexity, which manifests as duplicate reads (i.e., clonality) in the sequencing data (78, 79, 82); and the inevitably biased nature of the recovered library fragments (78, 79).

Table 1   Comparison of the three standard sequencing approaches used for ancient DNA, according to data and analyses characteristics

	Targeted SNP capture	Whole-genome capture	Whole-genome shotgun
Data characteristics			
Genomic coverage	Targeted SNPs and alleles	Genome-wide	Genome-wide
Typical enrichment range	45–13,000 (6)	2–13× (79)	None
Best use scenario	Low endogenous DNA; low/medium complexity	Low endogenous DNA; high complexity	Medium/high endogenous DNA; high complexity
Analyses characteristics[a]			
Diploid genotyping	Possible with high coverage, potential for allelic bias	Possible with high coverage, potential for allelic bias	Possible with high coverage
Ascertainment bias	Specific to targeted SNP panel	None	None
Suitability for merging with reference variant sets	Only variants overlapping with capture panel	All variants	All variants
Basic population structure and admixture analyses	Yes	Yes	Yes
Demographic inference	Methods not sensitive to ascertainment bias and/or allowing for correction	Yes	Yes
Rare variant analyses	Only variants overlapping with capture panel	Yes	Yes
Recovery of host-associated pathogens	Only if targeted with capture probes	Only if targeted with capture probes	Yes

[a]Analyses covered are typical examples for studies of species with population-level genomic data sets available (e.g., humans).

Abbreviation: SNP, single-nucleotide polymorphism.

In the case of environmental aDNA, amplicon-based sequencing (i.e., metabarcoding) is the primary low-cost alternative to shotgun sequencing (83, 84). Specific genes, which comprise enough DNA variability to enable identification of their composition within the targeted group and are either unique to one particular species or shared by a group of taxa, are PCR-amplified prior to sequencing (11). The key disadvantage of this approach is the relatively short length of the DNA stretches that can be successfully targeted (normally <200 base pairs), which limits the potential taxonomic identification between closely related taxa (85).

A typical computational workflow for handling high-throughput sequencing data from aDNA samples consists of the following steps: (*a*) trimming adaptor sequences from the reads, (*b*) collapsing them when matching pairs with significant overlap are identified, (*c*) filtering out reads with lengths shorter than 25–30 base pairs, and (*d*) aligning the remaining reads against reference genomes. Following this, alignments that are assigned with low-quality scores, as well as PCR duplicates, are removed, and reads are locally realigned around small insertions and deletions to improve overall genome quality (reviewed in 13). The open-source package PALEOMIX (86) is a computational pipeline that implements most of these procedures. The alignment of DNA reads is usually followed by aDNA authentication and contamination quantification. Nucleotide misincorporation, originating from miscoding lesions and created during PCR amplification, is the basis of these two procedures. For instance, during the blunt-end ligation step of the double-stranded DNA library preparation, cytosine deamination at 5′ overhangs results in greater C→T and G→A misincorporation rates toward the starts and the ends of the sequences, respectively (87). Software packages such as mapDamage (88, 89) or pmdtools (90) can test for the presence of such

patterns of nucleotide misincorporation. Contamination can be quantified using mitogenomes (5), X chromosome data (91), or autosomal data (5, 92). Additionally, estimating error rates and genome completion is important for limiting their effects in downstream analyses (reviewed in 13).

Finally, the mapped genome(s) can undergo genotype calling if the average genomic coverage is high (typically >15×). If a genome is of low coverage, it is recommendable to either obtain genotype likelihoods (93) or perform pseudohaploid sampling (5), to enable unbiased comparisons between low- and high-coverage data. The ancient genome(s) can then be used for downstream analyses to infer evolutionary histories. These analyses often include exploratory approaches, such as principal-component analysis (94) and latent class modeling (95), as well as more parameter-rich demographic estimation methods (e.g., 96, 97) (**Figure 2a**).

Ancient proteins. Standard extraction methods, targeting intact proteins ordinarily found in modern organisms, need to be modified to deal effectively with the extended fragmentation and chemical alteration of ancient proteins and to minimize the losses associated with each preparation step (**Figure 2b**). MS has become the key method for the analysis of ancient proteins, providing a reliable sequencing tool for high-throughput, confident identification of ancient proteins and proteomes.

Subsampling and extraction. Transferring of the protein residues in solution is dismissed in favor of direct digestion of the whole demineralized organic bone matrix pellet, enabling trypsin to access and digest the crosslinked portion of the substrate and solubilize tryptic peptides (98). An alternative approach, based on SDS denaturation, enables enhanced protein recovery from soft tissues preserved in ethanol (99) and oral microbiome proteins in dental calculus (29). This approach, however, is better suited for relatively well-preserved ancient samples, as the repeated buffer exchanges can cause extensive losses of fragmented protein residues. A less destructive approach has been recently developed, where most of the conventional preparation steps, i.e., reduction, alkylation, and digestion, are omitted and instead all the solubilized peptides are directly processed by MS analysis. This method has been used to solubilize archaeological proteins from enamel by HCl surface etching, creating sampling holes only several microns deep in modern and archaeological tissues (36).

Sequencing and computational analysis. There are two complementary proteomics approaches, referred to as top-down and bottom-up proteomics (100, 101). Top-down proteomics is the analysis of proteins in their undigested form (proteoforms) to preserve valuable information about post-translational modifications, isoforms, and proteolytic processing. However, this approach is still in its infancy and has yet to be applied to ancient proteins. Consequently, bottom-up proteomics is the approach ordinarily used for ancient protein sequence identification, where a protein mixture is digested to peptides by proteolytic enzymes (e.g., trypsin). The resulting peptide mixture can then be identified either by PMF, based on MS analysis, or by nano liquid chromatography coupled with tandem MS (nanoLC-MS/MS). In PMF, the obtained peptide masses are compared to the masses of peptides predicted to derive from known proteins by the same enzymatic digestion (102). PMF-based approaches are now widely adopted to taxonomically assign morphologically unidentifiable bone fragments from the Holocene and Pleistocene, even in batches of hundreds to thousands of samples (47, 103).

In contrast, MS/MS enables peptide sequencing from complex protein mixtures. Given that ancient proteomes are typically neither particularly complex nor abundant, analytical settings should be adapted to prioritize sensitivity over speed (104). Despite all the technological improvements

introduced so far, only a fraction of the available peptide sample actually enters the mass spectrometer as gas-phase ions (105). The inclusion of dimethyl sulfoxide as a supercharging reagent in liquid chromatography (LC) solvents can enhance peptide electrospray ionization, improving their identification sensitivity (106).

Dedicated algorithms match the masses of the product ions, generated after isolation and collision-induced fragmentation of the peptides analyzed, against theoretical MS/MS spectra generated as described above. Byonic (107) and Andromeda (108, 109) computer-based search engines are particularly suitable for ancient protein analysis as they allow the identification of unknown modifications and highly degraded protein sequences. De novo sequencing is a powerful new approach for ancient protein identification. It requires no prior protein sequence knowledge, allowing the identification of novel amino acid substitutions (single amino acid polymorphisms, SAPs) compared to existing protein sequence databases (20). A hybrid solution, error-tolerant search algorithms, do utilize protein sequence databases while allowing amino acid substitutions, similarly enabling the identification of (novel) SAPs. However, despite the tremendous effort of researchers to develop de novo and error-tolerant sequencing algorithms, e.g., Lutefisk (110) and PEAKS (111), complete peptide identification remains difficult owing to the gaps in the ion series recorded in MS/MS spectra (112, 113). The application of de novo or error-tolerant search algorithms therefore introduces biases in ancient peptide and protein identifications, particularly when the analyzed sample is phylogenetically distant from available reference sequences (113). New de novo sequencing algorithms, e.g., algorithms based on deep learning techniques such as DeepNovo (114), have the potential to significantly alter this aspect of ancient protein analysis.

Ancient lipids. The recovery, sampling, and determinations of lipid biomarkers in archaeological, palaeontological, and/or sedimentary materials are built on solid foundations of robust molecular concepts and are backed up by numerous experimental studies. The particular category of sample, and likely compound class present therein, defines the analytical protocols and specific instrumental techniques employed. These requirements are further defined by the need for quantitative, qualitative, and compound-specific isotopic information.

Subsampling and extraction. Solvent extraction is the principal method used to solubilize or extract lipids from various substrates. Given the range of compound classes present, i.e., hydrocarbons to carboxylic acids, solvent systems of intermediate polarity are often used. Whole fat, wax, resinous, and bituminous residues are directly dissolved in solvents, while subsamples of rock, sediments, soils, fossils, potsherds, bone or porous stone artefacts, coprolites, etc., are first crushed to a powder to open the matrix and increase extraction efficiencies. Total lipids (**Figure 2c**) are extracted using organic solvent mixtures, such as chloroform/methanol 2:1 (*v/v*) or dichloromethane/methanol 2:1 (*v/v*), by ultrasonication in disposable glass vials to avoid cross contamination. Where larger samples (tens of milligrams to grams) are available, lipids can be efficiently recovered using continuous Soxhlet extraction, accelerated solvent extraction, or microwave extraction. These approaches are readily applicable to soft tissues after air or freeze drying (63). Fractionations by column or solid-phase extraction chromatography are used to separate complex mixtures when trace lipid components are targeted (115). If more polar lipids, such as diacids, are targeted, alkaline or acid extraction procedures are used (116). This results in the loss of some compositional information preserved in intact ester lipids, notably triacylglycerols and wax esters (62). A range of other chemical cleavage reagents have been used either singly or sequentially to release covalently (sulphur, ether, ester, aromatic) bound lipid biomarkers from geopolymers (117). Following chemical cleavage reactions, derivatization, e.g., trimethylsilylation or esterification, is

generally required to cap protic sites prior to GC, GC-MS, and/or GC–combustion–isotope ratio MS (GC-C-IRMS) (62).

Instrumental techniques. Direct MS analysis without chromatographic separation can be useful for fast characterization of visible lipid residues, organic fossils, and organic macerals. This approach is most suitable when the amounts of sample are limited. However, without chromatographic separation, mass spectra generally comprise mixtures of compounds, and the increased complexity makes interpretation challenging (118 and references therein). Thus, the most effective analyses of lipid biomarkers use GC to achieve molecular separations (119). Identifications of important biomarkers present only at trace levels can be achieved by GC-MS operated in the selected ion monitoring mode to provide enhanced sensitivity and selectivity (e.g., 115). Where further selectivity or sensitivity is required, then GC-MS/MS can be used with selected reaction monitoring. Compounds that cannot be volatilized (i.e., high-molecular-weight and polar lipids, or polymerized resins/geopolymers) are inaccessible to GC analysis, in which case pyrolysis-GC-MS can be used to fingerprint the thermal cleavage products of geopolymers, fossils, polymeric resins, and ambers (e.g., 120). LC-MS has also been routinely used in the analysis of bacterial and archaeal tetraether lipid biomarkers in organic geochemistry (121).

Conveniently, GC can be coupled to isotope ratio mass spectrometry (122). With GC-C-IRMS, stable carbon isotope values can be determined for individual compounds identified by GC-MS. Linking molecular structures to their stable isotope values offers enhanced specificity compared to bulk IRMS analysis and provides information relating to carbon cycling processes, metabolic origins, and paleoecology (e.g., 123). In this way, different botanical origins of plant waxes [C_3 versus C_4 plants; (123)] can be distinguished based on $\delta^{13}C$ values of the high carbon number *n*-alkanes (C_{27} to C_{33}) derived from leaf waxes. Similarly, the origins of animal fats in archaeological pottery can be determined based on the $\delta^{13}C$ values of the two fatty acids palmitic acid ($C_{16:0}$) and stearic acid ($C_{18:0}$) (124). Another variant of the GC-C-IRMS technique allows determination of the compound-specific nitrogen isotopes of amino acids in proteins, notably collagen. Combined with carbon isotopes this allows enhanced specificity in paleodietary studies by exploiting the contrasting dietary origins and metabolic behaviors of different amino acids (125, 126). Finally, compound-specific hydrogen isotope values of lipids, notably fatty acids, have been used to link changes in human diet and animal exploitation to variations in precipitation (127).

APPLICATIONS IN EVOLUTIONARY BIOLOGY

Analyses of ancient biomolecules have led to some of the biggest breakthroughs in the field of evolutionary biology. The knowledge generated to date spans continents and epochs and includes all the major groups of organisms. Here we cover some of the most prominent examples.

Archaic Hominins

Ancient genomics has been central to furthering our understanding of human evolution after our divergence from archaic hominins, as well as the evolutionary consequences of human encounters with archaic hominin groups in the Late Pleistocene (128). The comparison of the Neanderthal genome with non-African individuals showed that the ancestors of all present-day non-African people contain around 2% Neanderthal DNA, indicating that they admixed with Neanderthals shortly after the dispersal of humans from Africa approximately 65–55 kyr ago (129). Sequencing the genome of another group of archaic hominins—the enigmatic Denisovans, so far known only from a finger bone and several teeth excavated in the Denisova Cave—has shown that interbreeding

occurred between them and the ancestors of present-day Oceanian peoples (129). This and other archaic introgression events may have helped modern humans to adapt to local environmental conditions, such as high altitudes in Tibet (130, 131), and contributed to a wide range of modern human phenotypic traits.

Modern Human Ancestry and Demography

Much of the aDNA work to date has focused on the evolution and global dispersal of anatomically modern humans (**Figure 3a**). One particularly fruitful research area has been focusing on the admixture history and on the migration routes that have given rise to the present-day patterns of human genetic diversity (reviewed in 132). For instance, the genome of the 24,000-year-old child remains from Mal'ta in south-central Siberia showed strong genetic affinities with both western Eurasians and Native Americans, suggesting a dual ancestry of the First Americans (133). Moreover, the genome of a 12,600-year-old individual from the Anzick culture revealed closer genetic ties to Native Americans than to Europeans (68), ruling out a cross-Atlantic European origin for the Paleo-Indian Clovis culture in North America. Recently, the genome of an 11,500-year-old individual from interior Alaska revealed that the First Americans derive from a single source population and that they were most likely established in Beringia as early as 20,000 years ago (134). These and similar studies (7, 135, 136) have collectively demonstrated that much of the human genetic diversity we see today was created by migration and admixture events during human (pre)history.

Species Extinctions

While the Quaternary Megafaunal Extinction is very well recorded in the paleontological record, its causes remain hotly debated, with climate change and human overkill as the two principal candidates. Ancient DNA has shed new light on this debate. Using ancient mitochondrial DNA from megafaunal fossil remains spanning the past 50,000 years, Lorenzen et al. (3) showed that while climate may have been the main extinction driver for some megafaunal species, e.g., the Eurasian musk ox and woolly rhinoceros, it was the combined effects of climate and anthropogenic activities that led to the demise of others, e.g., the Eurasian steppe bison and wild horse. A follow-up study by Cooper et al. (137), focusing on nuclear DNA regions, reported similar findings: abrupt warming events brought about by interglacial periods have caused repeated population-level turnovers and created metapopulation structures highly vulnerable to the subsequent human impact. Environmental DNA from sediments also provided valuable information relating to megafaunal extinctions. For instance, it revealed the exact timing of megafaunal extinctions in different geographic regions, such as the late survival of the mammoth and horse in interior Alaska (138). Furthermore, it helped to pinpoint likely environmental drivers of megafaunal extinctions, such as a shift in Arctic vegetation assemblages from dry steppe-tundra dominated by forbs to moist tundra dominated by woody plants and graminoids at the end of the last glacial period (84).

Animal and Plant Domestication and Exploitation

The domestication of animals and plants over the past 11,500 years has offered invaluable insights into positive selection and the associated environmental adaptations. Studies of aDNA investigated the genetic footprint of domestication in various taxa such as pigs (139), dogs (140), and chickens (141), though arguably it is the horse that has received most of the attention to date, partly owing to its relatively rich and continuous paleontological record. Recent work has shed new light on the origin and spread of horse domestication (142) and has revealed how this process imposed positive selection on a range of genes involved with cognition, physiology, and locomotion; led to
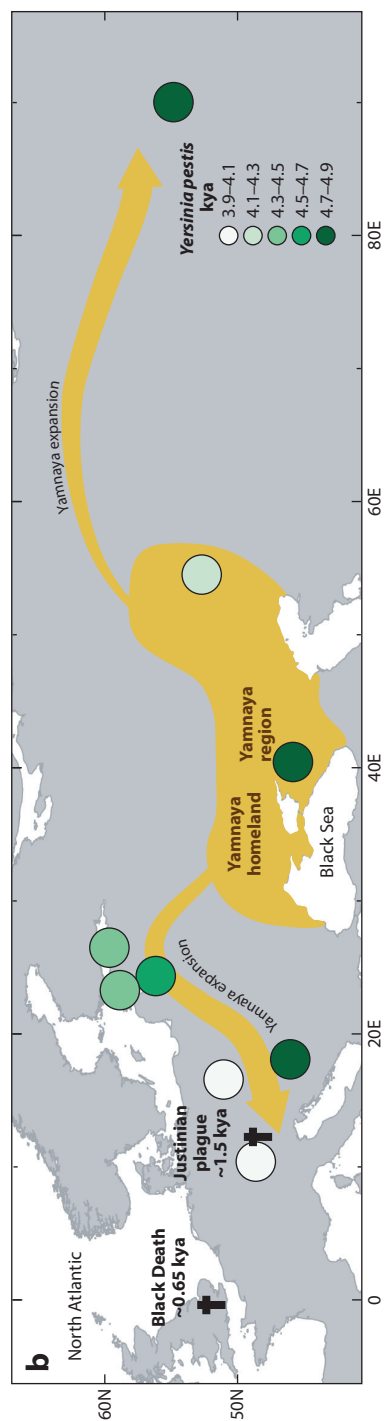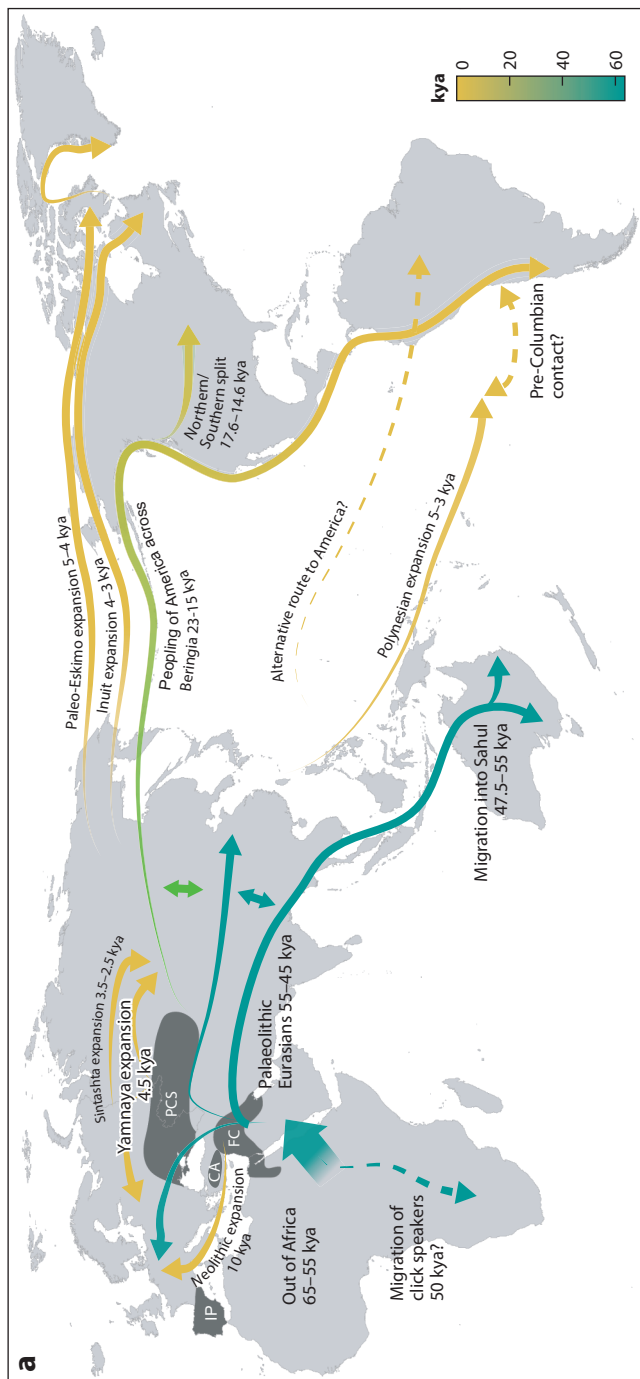
(Caption appears on following page)

Figure 3   (*Figure appears on preceding page*)

Major human migrations and spread of pathogens. (*a*) Migrations of modern humans as revealed by genomic data. Many of the depicted migration events were only possible to infer using ancient DNA data sets (e.g., Neolithic, Yamnaya, and Sintashta expansions in Western Eurasia) (132). (*b*) Spread of the Bronze Age plague. Circles indicate the geographic locations and the age of sites where genomic evidence of the plague-causing bacterium *Yersinia pestis* was isolated from ancient human remains (147). Chronology and geography correlate with the expansion (*arrows*) of peoples of the Yamnaya culture, as inferred from ancient genomes, suggesting diffusion of the disease through these prehistoric migrations. Abbreviation: CA, Central Anatolia; FC, Fertile Crescent; kya, thousands of years ago; IP, Iberian Peninsula; PCS, Pontic–Caspian steppe.

higher numbers of deleterious mutations; and resulted in a net loss of genetic diversity over the last two millennia (143, 144).

Plant domestication has received considerably less attention, possibly owing to the rarity of suitable plant macrofossil remains. Nevertheless, studies have revealed temporal patterns of adaptation and migration for important crops such as maize, wheat, cotton, and barley (reviewed in 31). This information may be used to facilitate the reintroduction of ancient alleles, lost at various stages of domestication, into modern animal breeds and plant varieties to advance breeding and management. Similarly, ancient biomolecules assisted with unraveling the beginnings of exploitation of wild species by humans. For instance, the analysis of lipid residues in more than 6,400 pottery vessels from across Europe, the Near East, and North Africa (145) identified beeswax in ceramic pottery sherds dating back as early as the seventh millennium BC, pointing to widespread exploitation of honeybee products by early farming societies in these regions.

Ancient Pathogens and Microbiomes

Shotgun sequencing of ancient skeletal remains can also reveal genetic information about the microorganisms originally associated with their host, from specific pathogens to entire microbiomes (**Figure 3***b*), while lipid analyses allow the detection of their cell wall biomarkers. Work to date has investigated the origin and evolution of some of the deadliest pathogens in human history, including the etiological agents of the Spanish flu and the bubonic plague—the H1N1 influenza virus (146) and *Yersinia pestis* (81), respectively. Interestingly, Rasmussen et al. (147) showed that bubonic plague had been widespread ∼3,000 years before any known written record and identified the temporal patterns of the appropriation of the genes contributing to the high pathogenicity of this bacterial species. The presence of cell wall biomarkers of *Mycobacterium tuberculosis*, namely mycolic acids, and the detection of *M. tuberculosis* aDNA in ancient individuals are seen as complementary evidence for ancient tuberculosis (148). Furthermore, deep sequencing of dental calculus and coprolites has identified bacteria typical of both oral and gut microbiomes in archaic, Neolithic, and medieval humans (29, 149, 150). These studies suggest that major dietary shifts in human history, such as neolithization and the Industrial Revolution, have caused a marked decrease in microbiome diversity and the rise of microbial taxa linked to chronic diseases. Future genomic studies of ancient microorganisms promise to bring major advances to our understanding of the evolution of human health and disease.

Phylogenetics of Extinct Taxa

Protein sequence variation has proven to be large enough to allow phylogenetic inference for extinct species recalcitrant to aDNA analysis. For instance, COL1 was used to elucidate the phylogeny of *Macraucbenia patachonica* and *Toxodon platensis*, two species of South American native ungulates, a taxonomic group of ∼280 recently extinct placental mammal species (20, 151). These studies placed these two species as a sister-clade to perissodactyls. The conclusions of these studies

were confirmed by aDNA analysis in a more recent study by Westbury et al. (152). Similarly, Cleland et al. (153) used collagen sequences to place the species *Castoroides ohioensis*, a Pleistocene giant beaver, within the Rodentia group. Prior to this, Buckley (154) sequenced ancient collagen from the so-called Malagasy aardvark (*Plesiorycteropus*), finding it more closely related to tenrecs than to aardvarks. Recent research indicates that the analysis of ancient proteins is a particularly promising approach to studying phylogenetics of Late Pleistocene hominins (e.g., Neanderthals, Denisovans), especially for regions and time periods in which ancient nuclear DNA is unlikely to survive (47).

## Prehistoric Milk Use and the Evolution of Lactase Persistence

The widespread preservation of animal fats in archaeological pottery results from animal product processing, with the saturated fatty acids, palmitic ($C_{16:0}$) and stearic ($C_{18:0}$), dominating archaeological lipid assemblages (119). Compound-specific carbon isotope analysis of the latter two compounds enables discrimination between fats from ruminants and from nonruminants, as well as between fats from the carcass and from the milk of ruminants, as shown in **Figure 4** (124). Evershed and coworkers (155) examined more than 2,200 pottery vessels from sites in the Near East and southeastern Europe to provide the earliest evidence for milk processing in the seventh millennium BC. This finding raises questions about whether lactase persistence, i.e., the postweaning ability to digest lactose, was required for the development of early dairy economies. These prehistoric farmers may well have been processing milk in pots to reduce the lactose to make their dairy products more digestible. Indeed, investigations of pottery from a cattle-herding site in central Poland dating to the sixth millennium BC demonstrated specialized milk processing using sieves, pointing to cheese-making (156). Studies of aDNA showed that the lactase persistence allele—located in a regulatory region upstream of *LCT*—rose to high frequency in Europe only quite recently, in the last 4,000 years (157), and may have been introduced initially into northern Europe by steppe populations migrating westward (7). The detection of horse milk in pottery from the Eurasian steppe using compound-specific carbon and hydrogen isotope values of fatty acids in tandem, together with skeletal morphologies, further supports the aDNA evidence (127).

Evolution of Grasses

The modern global ecosystem has a significant component of C_4 plants, comprising drought-adapted grasses and sedges that mostly grow in tropical savannas, temperate grasslands, and semidesert shrublands. Yet, while we know that C_4 carbon fixation is the youngest and most advanced of the three photosynthetic pathways used by plants, we are less certain when and where C_4 plants originated and how they expanded across the planet. Cerling and coworkers (159) addressed the question of the global expansion of C_4 plants by measuring the carbon isotope composition in mammalian tooth enamel and paleosol carbonate from Pakistan and North America across the late Miocene. The results showed a shift of carbon isotope values of 10‰ at approximately 7–5 million years ago (mya) both in the New and the Old World, which was interpreted as a global expansion of C_4 plants. Following up on this, Freeman & Colarusso (160) studied Siwalk paleosols from Pakistan, Nepal, and Bengal Fan sediments for the abundances and carbon isotopic composition of plant cuticular wax *n*-alkanes to examine molecular evidence for the expansion of C_4 grasses on the Indian subcontinent. They showed that the carbon isotopic values of high-carbon-number *n*-alkanes (C_{27} to C_{33}), both in the ancient soils and ocean sediments, experienced a shift from low $\delta^{13}C$ values (approximately −30‰) to higher values (approximately −22‰) prior to 6 mya, providing further support to the Miocene C_4 plant expansion hypothesis.

Origin of Early Life Forms

Owing to their superior preservation over longer time periods, lipid biomarkers have played a key role in determining the timing of major evolutionary innovations in early life forms. For instance, Brocks and coworkers examined lipid biomarkers in the Pilbara Craton shales (Australia) in order to identify the types of microorganisms inhabiting the oceans in Archaean times >2,000 mya (161). Among the complex lipid distributions, they were able to discern hopane biomarkers. The presence of 2α-methyl hopanes, diagnostic of cyanobacteria, helped constrain the age of the oldest
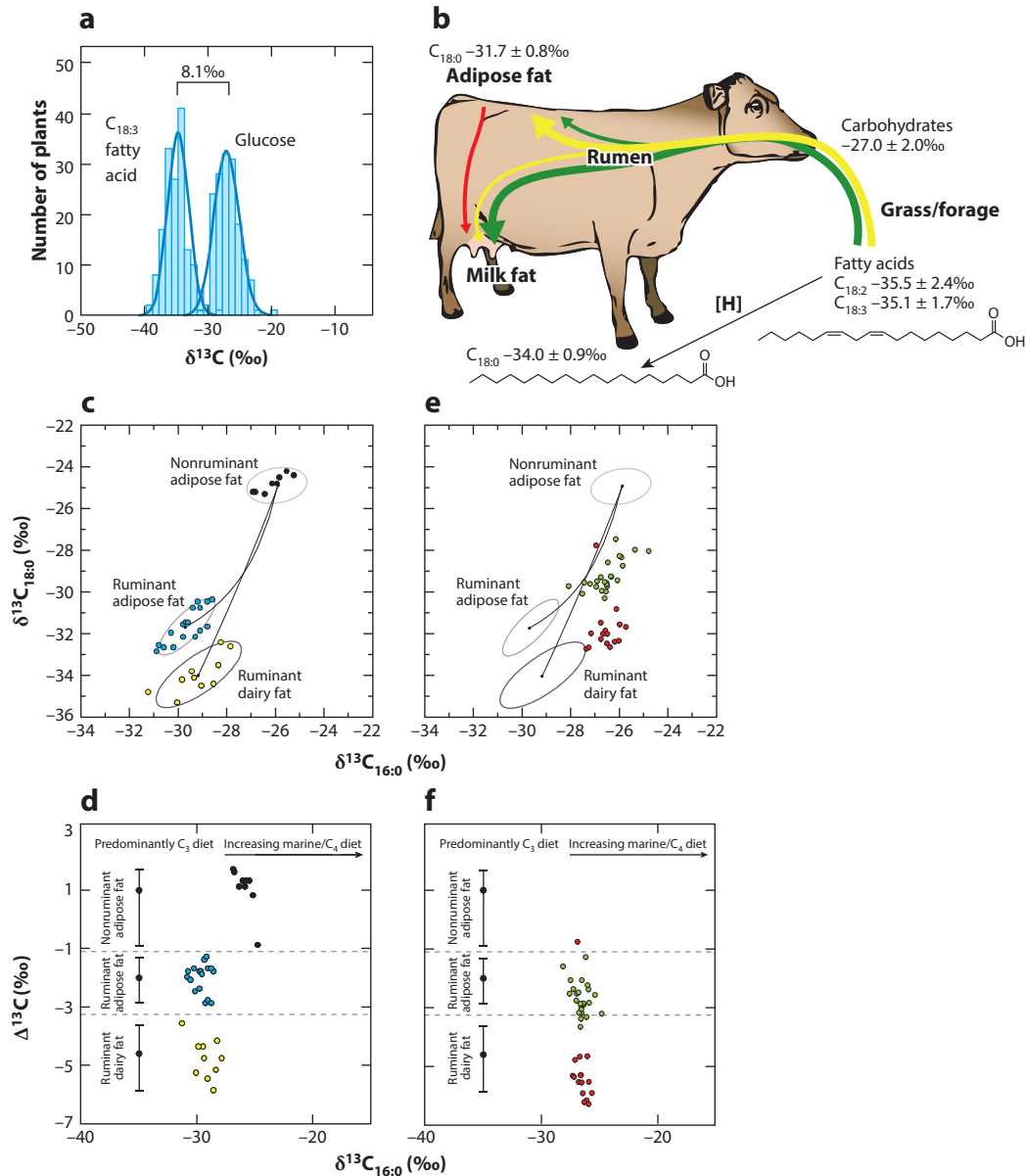
(Caption appears on following page)

Figure 4 (*Figure appears on preceding page*)

Basis of use of compound-specific isotope values of fatty acids to distinguish ancient animal fats. (*a*) Histogram of the $\delta^{13}C$ values of $C_{18:3}$ fatty acid and glucose extracted from plants. The histogram demonstrates an 8.1‰ mean difference in the $\delta^{13}C$ values of $C_{18:3}$ fatty acid (mean $= -36.3‰$) and glucose (mean $= -28.2‰$), and these isotopic differences are known to result from fractionation during the formation of acetylCoA. (*b*) Diagram showing the routing of dietary fatty acids and carbohydrates in the rumen, adipose tissue, and mammary gland of the ruminant animal. Approximately 60% of the $C_{18:0}$ in ruminant milk appears to be directly incorporated from the diet after biohydrogenation of unsaturated fatty acids (e.g., $C_{18:3}$) in the rumen and reflects the inability of the mammary gland to biosynthesize $C_{18:0}$. The difference in the $\delta^{13}C$ values of $C_{18:0}$ in ruminant adipose tissues and dairy fats can also be seen in the graphs in panels *c* and *d*. (*c*) Plot of the $\delta^{13}C$ values of the $C_{18:0}$ and $C_{16:0}$ fatty acids obtained from modern reference fats, with $p = 0.684$ confidence ellipses. (*d*) Plot of the $\Delta^{13}C$ ($= \delta^{13}C_{18:0} - \Delta^{13}C_{16:0}$) values of the major fatty acid components ($C_{16:0}$ and $C_{18:0}$) of modern reference fats, demonstrating that animal fats are also distinguished by using this criterion. The three fields correspond to the ranges of $\delta^{13}C$ values of the domesticates known to comprise the major component of prehistoric economies. All of the animals were raised on C_3 diets in Britain. The $\delta^{13}C$ values obtained from the modern reference materials have been adjusted for post–Industrial Revolution effects of fossil fuel burning by the addition of 1.2‰. (*e, f*) Example of an application of the $\delta^{13}C$ and $\Delta^{13}C$ proxies as a means to identify the source of animal fats in archaeological pottery sherds. Plots of the $\delta^{13}C$ values for the $C_{16:0}$ and $C_{18:0}$ fatty acids (*e*) and $\Delta^{13}C$ values (*f*) prepared from animal fat residues extracted from sherds of Neolithic sieves (*red*) and cooking pots (*green*) from the region of Kuyavia (Poland). Ellipses and ranges are those for modern animal fats depicted in panels *c* and *d*. Each data point represents an individual vessel. The study concluded that sieves were used for processing milk products to make cheese, while cooking pots were mainly used to process ruminant meat products. Note that the differences in the $\delta^{13}C$ values between modern and archaeological animal fats (*c, e*) are due to environmental factors (e.g., diet) and are removed when using the $\Delta^{13}C$ values (*d, f*). Analytical precision is $\pm 0.3‰$. Panels *a–d* adapted from Reference 158.

cyanobacteria and suggested that oxygenic photosynthesis had evolved well before the atmosphere became oxidizing (162). Similarly, a fossil molecular marker diagnostic for demosponges, 24-isopropylcholestane, has provided the oldest evidence for animals in the fossil record found to date (163). This molecular marker was found in Oman Salt Basin, Siberia, in Precambrian marine sequences older than 635 myr, which is long before the spicules of demosponges appear in the fossil record. These findings demonstrate the importance of chemical fossils for studying early life forms, particularly in regions and rock types where body fossils are rare.

Reconstructing Environmental Drivers of Evolution

Lipid biomarkers have a central role to play in determining past climatic conditions, such as temperature and precipitation, which would have been critical drivers of evolutionary changes. For example, ocean sediments offer two major proxies for reconstructing past temperatures. First, the alkenones biomarkers deriving from the planktonic alga *Emiliania huxleyi* were detected in ocean sediments and the degree of unsaturation in the C_{37} homologs was found to vary linearly with sea surface temperature, covarying with $\delta^{18}O$ values of foramiferal carbonate (164). The proxy, termed U^k_{37}, has been rigorously calibrated in laboratory cultures and through analyses of suspended particulates, core top sediments, and in situ measurements of sea surface temperature. As a result, U^k_{37} can now provide paleotemperature estimates going back at least 2.5 myr with an accuracy of $\pm 1°C$. The second sea surface temperature proxy, TEX_{86}, is based on glycerol dialkyl glycerol tetraethers (GDGTs), which serve as membrane lipids in Archaea (121). Though GDGTs provide temperature estimates (based on variations in the number of cyclopentyl and cyclohexyl rings and/or methyl branches in their structures) of lower precision compared to U^k_{37} ($\pm 2.5°C$), the former survive in sediments for tens to hundreds of millions of years. A notable example of the power of these molecular paleothermometers was in the investigation of global warming 55 mya at the Paleocene/Eocene Thermal Maximum. The GDGT biomarkers from a core drilled near the North Pole predicted mean annual sea surface temperatures of $>20°C$, and thus, an ice-free world (165).

FUTURE DIRECTIONS

The field of ancient biomolecules is likely to take numerous new directions over the coming years. Here we outline a selected few that are likely to make significant advances with respect to evolutionary inference.

Large-Scale, High-Coverage Genome Panels

Large-scale ancient genomic projects have so far relied either on low-coverage genome sequences (∼1X average coverage) or targeted capture of common genomic variants, owing to the prohibitive costs of generating high-coverage ancient genomic data (13). The number of individuals investigated in a single study is now comparable to those profiled in studies on living populations (166). Although these approaches are well suited for the analysis of deep demographic history at broader, continental-level scales, they pose severe limitations on the ability to infer more recent demographic events and separate more closely related populations. Thus, an important next step in aDNA research will be to routinely sequence large numbers of high-coverage genomes spanning larger spatial and temporal scales. This progress will be essential for facilitating the reconstruction of ancient epigenomes, elucidating selective processes, pinpointing the timing of divergence and admixture events, robustly detecting past inbreeding, and determining detailed family relationships between individuals. Large-scale, high-coverage genome panels will thus be instrumental in shifting the research focus from pattern description to mechanistic explanation of evolutionary processes.

High-Throughput Sequencing of Single Microfossils

Presently, the reconstruction of past population dynamics is limited to taxa with relatively continuous macrofossil records. This excludes most unicellular, as well as some major multicellular, groups of organisms (e.g., bacteria, fungi, plants), whose paleorecord is dominated by microfossils (e.g., pollen, fungal and bacterial spores). We therefore anticipate that a major new research direction in the field will be the deployment of high-throughput sequencing of individual microfossils aided by the rise of novel sequencing technologies. Single-cell sequencing technologies, which allow genome-wide sequencing of hundreds to thousands of single cells in parallel at a relatively low cost (167), seem to be a particularly promising research direction. If successful, high-throughput microfossil sequencing will enable simultaneous retrieval of a large number of individual genomes belonging to several different groups of organisms and thereby enable unprecedentedly detailed investigation into the evolutionary history of hundreds to thousands of species in parallel.

Deep-Time Phylogenetics

Ancient proteins can survive considerably longer than aDNA (61). Consequently, paleoproteomics has the potential to provide access to genetic evidence from epochs and geographic areas incompatible with aDNA preservation and enable investigation into deep-time evolution, which has so far been intractable for molecular phylogenetics. For this purpose, confident sequence reconstruction and coverage maximization should be prioritized over identification of a high number of generally poorly covered proteins. We expect that paleoproteomic analysis will increasingly rely both on stringent probability scores and confident sequence reconstruction through either a nearly complete fragment ion series or multiple coverage of the same amino acid position in partially overlapping peptides (113). Although COL1 is one of the most abundant and stable proteins,

often surviving for millions of years in biomineralized tissues, its value for phylogenetic inference can be inadequate owing to its limited variability (48). For example, the most recent phylogenetic analysis based on collagen sequences from a nonavian dinosaur, *Brachylophosaurus canadensis*, was unstable depending on the subsets of peptides examined (168). We therefore call for further efforts to identify proteins that possess enough variability to permit more detailed inferences about evolutionary processes beyond the last one million years (169, 170).

Ancient proteomes could be used to resolve the systematics of long-extinct groups of organisms, especially in instances where there are high levels of convergence among distantly related taxa. Condylarths, an informal group of extinct primitive ungulates, could be an ideal candidate target for this new approach. This label is largely used to classify ungulates that have not been clearly assigned to either Perissodactyla or Cetartiodactyla and is most probably composed of several unrelated mammalian lineages (171). Future studies focusing on condylarths and other similar groups, while following the above-discussed principles, may facilitate significant alteration of early mammalian radiations.

Recently, a study by Vajda et al. (172) has revealed the great promise of ancient leaf waxes for understanding the phylogenies of extinct plant taxa whose fossils do not yield aDNA sequences. The authors analyzed geothermally resistant fossil cuticles of seed plants using Fourier transform infrared spectroscopy. These cuticles are composed of an insoluble membrane of lipids and hydrocarbon polymers impregnated with soluble waxes that can survive on very long timescales, even as far back as the earliest appearance of land plants. The results showed that the cuticles can preserve biomolecular information that can consistently differentiate major taxa. This holds true for cuticles with markedly different diagenetic histories, highlighting the potential universality of this proxy. Considering that fossil leaves are available for a wide range of extinct plant taxa, this new proxy offers an exciting opportunity to reconstruct extinct plant relationships far into the geological past.

Evolution of Protein Structure and Function

On a longer perspective, the robust reconstruction of very ancient protein sequences has the potential to help advance theoretical elements in molecular evolution studies, such as ancestral amino acid sequence reconstruction (ASR) (173, 174). It is still not fully understood whether the methods currently used for this purpose are reliable, neither is it clear which of the available methods should be chosen. Furthermore, current ASR studies do not incorporate protein sequences derived from extinct clades, limiting ASR inferences to the phylogenetic processes behind the protein evolution of extant species. Thus, by using ancient protein sequencing, we may also be able to advance our understanding of protein evolution by directly observing ancient protein sequences at critical evolutionary moments, enabling the reconstruction of the processes behind losses or acquisitions of function, with potential implications for future bioengineering of novel proteins. Simultaneously, approaches such as ASR, structural modeling, and the functional analysis of (reconstructed ancient) proteins might provide insights into the driving forces behind protein sequence evolution in now-extinct clades (175).

Ancient Lipidomics

The study of lipidomes investigates whole lipid assemblages as opposed to individual compounds, relying on newly developed high-resolution instrumentation and dedicated data-processing workflows. The production of large data sets—large both in terms of numbers of individual compounds, which can be in the thousands in biological samples, and numbers of individual samples

analyzed—is highly challenging but opens up new exciting avenues of investigation. For example, lipidomics can reveal low-abundance compounds to be of high relevance for elucidating biochemical or biosynthetic links between compounds or prompt biomarker discoveries (176). Although further development of multiparametrical statistical approaches will be essential to overcome the current challenges of effectively analyzing huge data volumes, this field holds great potential for advancing our understanding of the origins of lipid signatures in the geological record and the establishment of novel lipid biomarkers.

Multiproxy Approaches

The field of ancient biomolecules has now entered a new phase, embarking on large-scale, diachronic analyses that provide broader perspectives on the origin and evolution of life across temporal and spatial spheres. Recent studies have shown how the use of a combination of ancient biomolecules can provide a more coherent approach to tackling complex biological questions than the use of any of these biomolecular proxies alone. For instance, Warinner et al. (29) have combined the analyses of aDNA and proteins to provide high-resolution taxonomic and functional characterization of the ancient oral microbiome that enabled simultaneous investigation of pathogen activity, host immunity, and diet in past populations. We anticipate that future studies will increasingly adopt such multiproxy approaches, integrating the analyses of different ancient biomolecules and combining them with environmental and cultural proxies in order to provide unprecedented insights into the evolutionary history of life on Earth.

CONCLUSION

Studies of ancient biomolecules have come a long way—from retrieval of short sequences of mitochondrial DNA from late Holocene materials to the assembly of full nuclear genome sequences and characterization of proteins and lipids dating back millions of years. These studies have profoundly deepened our understanding of the origin of early life forms, adaptation and extinction processes, and past migrations and admixtures that gave rise to present-day biological diversity, including in our own species. Today, ancient biomolecules can provide direct insights into both the deep and recent evolutionary past, at a scale and level of detail that few would have predicted less than a decade ago.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

LITERATURE CITED

1. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312(5991):282–84
2. Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, et al. 2004. Rise and fall of the Beringian steppe bison. *Science* 306(5701):1561–65
3. Lorenzen ED, Nogués-Bravo D, Orlando L, Weinstock J, Binladen J, et al. 2011. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* 479(7373):359–64
4. Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282):757–62
5. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328(5979):710–22
6. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522(7555):207–11
7. Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522(7555):167–72
8. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, et al. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499(7456):74–78
9. Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, et al. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300(5620):791–95
10. Willerslev E, Cappellini E, Boomsma W, Nielsen R, Hebsgaard MB, et al. 2007. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* 317(5834):111–14
11. Pedersen MW, Overballe-Petersen S, Ermini L, Sarkissian CD, Haile J, et al. 2015. Ancient and modern environmental DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370(1660):20130383
12. Pedersen MW, Ruter A, Schweger C, Friebe H, Staff RA, et al. 2016. Postglacial viability and colonization in North America's ice-free corridor. *Nature* 537(7618):45–49
13. Orlando L, Gilbert MTP, Willerslev E. 2015. Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* 16(7):395–408
14. Abelson PH. 1954. Amino acids in fossils. *Science* 119(3096):576
15. Lowenstein JM. 1981. Immunological reactions from fossil material. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2921057:143–49
16. Huq NL, Tseng A, Chapman GE. 1990. Partial amino acid sequence of osteocalcin from an extinct species of ratite bird. *Biochem. Int.* 21(3):491–96
17. Ostrom PH, Schall M, Gandhi H, Shen T-L, Hauschka PV, et al. 2000. New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry. *Geochim. Cosmochim. Acta* 64(6):1043–50
18. Buckley M, Collins M, Thomas-Oates J, Wilson JC. 2009. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 23(23):3843–54
19. Cappellini E, Jensen LJ, Szklarczyk D, Ginolhac A, da Fonseca RAR, et al. 2012. Proteomic analysis of a Pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *J. Proteome Res.* 11(2):917–26
20. Welker F, Collins MJ, Thomas JA, Wadsley M, Brace S, et al. 2015. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature* 522(7554):81–84
21. Bergmann W. 1963. Geochemistry of lipids. In *Organic Geochemistry*, ed. IA Breger, pp. 503–42. Oxford, UK: Pergamon
22. Thornton MD, Morgan ED, Celoria F. 1970. The composition of bog butter. *Sci. Archaeol.* 1(2 3):20–25
23. Killops SD, Killops VJ. 1993. *An Introduction to Organic Geochemistry*. Harlow, UK: Longman. 265 pp.

24. Briggs DEG, Evershed RP, Lockheart MJ. 2000. The biomolecular paleontology of continental fossils. *Paleobiology* 26(sp4):169–93

25. Evershed RP. 2008. Organic residue analysis in archaeology: the archaeological biomarker revolution. *Archaeometry* 50(6):895–924

26. Summons RE, Powell TG, Boreham CJ. 1988. Petroleum geology and geochemistry of the Middle Proterozoic McArthur Basin, Northern Australia: III. Composition of extractable hydrocarbons. *Geochim. Cosmochim. Acta* 52(7):1747–63

27. Dunne J, Evershed RP, Salque M, Cramp L, Bruni S, et al. 2012. First dairying in green Saharan Africa in the fifth millennium BC. *Nature* 486(7403):390–94

28. Green EJ, Speller CF. 2017. Novel substrates as sources of ancient DNA: prospects and hurdles. *Genes* 8(7):180

29. Warinner C, Rodrigues JFM, Vyas R, Trachsel C, Shved N, et al. 2014. Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* 46(4):336–44

30. Gilbert MTP, Tomsho LP, Rendulic S, Packard M, Drautz DI, et al. 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* 317(5846):1927–30

31. Wales N, Allaby R, Willerslev E, Gilbert MTP. 2013. Ancient plant DNA. In *Encyclopedia of Quaternary Science*, ed. SA Elias, CJ Mock, pp. 705–15. Amsterdam: Elsevier. 2nd ed.

32. Slon V, Hopfe C, Weiß CL, Mafessoni F, de la Rasilla M, et al. 2017. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* 356(6338):605–8

33. Bennett KD, Parducci L. 2006. DNA from pollen: principles and potential. *Holocene* 16(8):1031–34

34. Shoulders MD, Raines RT. 2009. Collagen structure and stability. *Annu. Rev. Biochem.* 78:929–58

35. Salmon CR, Tomazela DM, Ruiz KGS, Foster BL, Paes Leme AF, et al. 2013. Proteomic analysis of human dental cementum and alveolar bone. *J. Proteom.* 91:544–55

36. Stewart NA, Gerlach RF, Gowland RL, Gron KJ, Montgomery J. 2017. Sex determination of human remains from peptides in tooth enamel. *PNAS* 114(52):13649–54

37. Dallongeville S, Garnier N, Rolando C, Tokarski C. 2016. Proteins in art, archaeology, and paleontology: from detection to identification. *Chem. Rev.* 116(1):2–79

38. Fiddyment S, Holsinger B, Ruzzier C, Devine A, Binois A, et al. 2015. Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting. *PNAS* 112(49):15066–71

39. Cappellini E, Gilbert MTP, Geuna F, Fiorentino G, Hall A, et al. 2010. A multidisciplinary study of archaeological grape seeds. *Naturwissenschaften* 97(2):205–17

40. Clark KA, Ikram S, Evershed RP. 2013. Organic chemistry of balms used in the preparation of pharaonic meat mummies. *PNAS* 110(51):20392–95

41. Charters S, Evershed RP, Goad LJ, Leyden A, Blinkhorn PW, Denham V. 1993. Quantification and distribution of lipid in archaeological ceramics: implications for sampling potsherds for organic residue analysis and the classification of vessel use. *Archaeometry* 35(2):211–23

42. Berstan R, Dudd SN, Copley MS, Morgan ED, Quye A, Evershed RP. 2004. Characterisation of "bog butter" using a combination of molecular and isotopic techniques. *Analyst* 129(3):270–75

43. Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, et al. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. Biol. Sci.* 279(1748):4724–33

44. Kistler L, Ware R, Smith O, Collins M, Allaby RG. 2017. A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Res.* 45(11):6310–20

45. Hansen A, Willerslev E, Wiuf C, Mourier T, Arctander P. 2001. Statistical evidence for miscoding lesions in ancient DNA templates. *Mol. Biol. Evol.* 18(2):262–65

46. Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S. 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* 29(23):4793–99

47. Welker F, Hajdinjak M, Talamo S, Jaouen K, Dannemann M, et al. 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *PNAS* 113(40):11162–67

48. Welker F, Smith GM, Hutson JM, Kindler L, Garcia-Moreno A, et al. 2017. Middle Pleistocene protein sequences from the rhinoceros genus Stephanorhinus and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae. *PeerJ* 5:e3033

49. Mackenzie AS, Brassell SC, Eglinton G, Maxwell JR. 1982. Chemical fossils: the geological fate of steroids. *Science* 217(4559):491–504
50. Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362(6422):709–15
51. Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, et al. 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *PNAS* 109(36):E2382–90
52. Meyer M, Arsuaga J-L, de Filippo C, Nagel S, Aximu-Petri A, et al. 2016. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* 531(7595):504–7
53. Bremond L, Favier C, Ficetola GF, Tossou MG, Akouégninou A, et al. 2017. Five thousand years of tropical lake sediment DNA records from Benin. *Quat. Sci. Rev.* 170:203–11
54. Geiger T, Clarke S. 1987. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J. Biol. Chem.* 262(2):785–94
55. Robinson AB, Rudd CJ. 1974. Deamidation of glutaminyl and asparaginyl residues in peptides and proteins. *Curr. Top. Cell. Regul.* 8:247–95
56. Joshi AB, Kirsch LE. 2004. The estimation of glutaminyl deamidation and aspartyl cleavage rates in glucagon. *Int. J. Pharm.* 273(1–2):213–19
57. van Doorn NL, Wilson J, Hollund H, Soressi M, Collins MJ. 2012. Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun. Mass Spectrom.* 26(19):2319–27
58. Cleland TP, Schroeter ER, Schweitzer MH. 2015. Biologically and diagenetically derived peptide modifications in moa collagens. *Proc. Biol. Sci.* 282(1808):20150015
59. Schweitzer MH, Zheng W, Cleland TP, Goodwin MB, Boatman E, et al. 2014. A role for iron and oxygen chemistry in preserving soft tissues, cells and molecules from deep time. *Proc. Biol. Sci.* 281(1775):20132741
60. Hill RC, Wither MJ, Nemkov T, Barrett A, D'Alessandro A, et al. 2015. Preserved proteins from extinct *Bison latifrons* identified by tandem mass spectrometry; hydroxylysine glycosides are a common feature of ancient collagen. *Mol. Cell. Proteom.* 14(7):1946–58
61. Demarchi B, Hall S, Roncal-Herrero T, Freeman CL, Woolley J, et al. 2016. Protein sequences bound to mineral surfaces persist into deep time. *eLife* 5:e17092
62. Correa-Ascencio M, Evershed RP. 2014. High throughput screening of organic residues in archaeological potsherds using direct acidified methanol extraction. *Anal. Methods* 6(5):1330
63. Corr LT, Richards MP, Jim S, Ambrose SH, Mackie A, et al. 2008. Probing dietary change of the Kwäday Dän Ts'ìnchi individual, an ancient glacier body from British Columbia: I. Complementary use of marine lipid biomarker and carbon isotope signatures as novel indicators of a marine diet. *J. Archaeol. Sci.* 35(8):2102–10
64. Evershed RP, Bland HA, van Bergen PF, Carter JF, Horton MC, Rowley-Conwy PA. 1997. Volatile compounds in archaeological plant remains and the Maillard reaction during decay of organic matter. *Science* 278(5337):432–33
65. Evershed RP. 1993. Biomolecular archaeology and lipids. *World Archaeol.* 25(1):74–93
66. Evershed RP. 2008. Experimental approaches to the interpretation of absorbed organic residues in archaeological ceramics. *World Archaeol.* 40(1):26–47
67. Craig OE, Saul H, Lucquin A, Nishida Y, Taché K, et al. 2013. Earliest evidence for the use of pottery. *Nature* 496(7445):351
68. Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, et al. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506(7487):225–29
69. Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, et al. 2015. Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* 5:11184
70. Sirak KA, Fernandes DM, Cheronet O, Novak M, Gamarra B, et al. 2017. A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *Biotechniques* 62(6):283–89
71. Romanowski G, Lorenz MG, Wackernagel W. 1991. Adsorption of plasmid DNA to mineral surfaces and protection against DNase I. *Appl. Environ. Microbiol.* 57(4):1057–61
72. Wales N, Andersen K, Cappellini E, Avila-Arcos MC, Gilbert MTP. 2014. Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLOS ONE* 9(1):e86827

73. Malmstrom H, Svensson EM, Gilbert MTP, Willerslev E, Gotherstrom A, Holmlund G. 2007. More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. *Mol. Biol. Evol.* 24(4):998–1004

74. Barnett R, Larson G. 2012. A phenol-chloroform protocol for extracting DNA from ancient samples. *Methods Mol. Biol.* 840:13–19

75. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, et al. 2007. Direct selection of human genomic loci by microarray hybridization. *Nat. Methods* 4(11):903–5

76. Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, et al. 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* 328(5979):723–25

77. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, et al. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* 27(2):182–89

78. Cruz-Dávalos DI, Llamas B, Gaunitz C, Fages A, Gamba C, et al. 2017. Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol. Ecol. Resour.* 17(3):508–22

79. Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, et al. 2013. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am. J. Hum. Genet.* 93(5):852–64

80. Vinner L, Mourier T, Friis-Nielsen J, Gniadecki R, Dybkaer K, et al. 2015. Investigation of human cancers for retrovirus by low-stringency target enrichment and high-throughput sequencing. *Sci. Rep.* 5:13201

81. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, et al. 2011. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478(7370):506–10

82. Avila-Arcos MC, Cappellini E, Romero-Navarro JA, Wales N, Moreno-Mayar JV, et al. 2011. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci. Rep.* 1:74

83. Pedersen MW, Ginolhac A, Orlando L, Olsen J, Andersen K, et al. 2013. A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quat. Sci. Rev.* 75:161–68

84. Willerslev E, Davison J, Moora M, Zobel M, Coissac E, et al. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* 506(7486):47–51

85. Coissac E, Riaz T, Puillandre N. 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Mol. Ecol.* 21(8):1834–47

86. Schubert M, Ermini L, Der Sarkissian C, Jónsson H, Ginolhac A, et al. 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* 9(5):1056–82

87. Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KAS, et al. 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC Genom.* 13:178

88. Orlando L, Ginolhac A, Raghavan M, Vilstrup J, Rasmussen M, et al. 2011. True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Res.* 21(10):1705–19

89. Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* 27(15):2153–55

90. Skoglund P, Northoff BH, Shunkov MV, Derevianko AP, Pääbo S, et al. 2014. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *PNAS* 111(6):2229–34

91. Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, et al. 2011. An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334(6052):94–98

92. Racimo F, Renaud G, Slatkin M. 2016. Joint estimation of contamination, error and demography for nuclear DNA from ancient humans. *PLOS Genet.* 12(4):e1005972

93. Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformat.* 15:356

94. Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLOS Genet.* 2(12):e190

95. Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–59

96. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLOS Genet.* 9(10):e1003905

97. Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–96

98. Cleland TP. 2018. Solid digestion of demineralized bone as a method to access potentially insoluble proteins and post-translational modifications. *J. Proteome Res.* 17(1):536–42

99. Cappellini E, Gentry A, Palkopoulou E, Ishida Y, Cram D, et al. 2014. Resolution of the type material of the Asian elephant, *Elephas maximus* Linnaeus, 1758 (Proboscidea, Elephantidae). *Zool. J. Linn. Soc.* 170(1):222–32

100. Fornelli L, Durbin KR, Fellers RT, Early BP, Greer JB, et al. 2017. Advancing top-down analysis of the human proteome using a benchtop quadrupole-orbitrap mass spectrometer. *J. Proteome Res.* 16(2):609–18

101. Toby TK, Fornelli L, Kelleher NL. 2016. Progress in top-down proteomics and the analysis of proteo-forms. *Annu. Rev. Anal. Chem.* 9(1):499–519

102. Aebersold R, Goodlett DR. 2001. Mass spectrometry in proteomics. *Chem. Rev.* 101(2):269–95

103. Brown S, Higham T, Slon V, Pääbo S, Meyer M, et al. 2016. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci. Rep.* 6:23559

104. Kelstrup CD, Young C, Lavallee R, Nielsen ML, Olsen JV. 2012. Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. *J. Proteome Res.* 11(6):3487–97

105. Wilm M, Mann M. 1996. Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* 68(1):1–8

106. Hahne H, Pachl F, Ruprecht B, Maier SK, Klaeger S, et al. 2013. DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat. Methods* 10(10):989–91

107. Bern M, Kil YJ, Becker C. 2012. Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinform.* 40:13.20.1–13.20.14

108. Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26(12):1367–72

109. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10(4):1794–805

110. Taylor JA, Johnson RS. 1997. Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 11(9):1067–75

111. Ma B, Zhang K, Hendrie C, Liang C, Li M, et al. 2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 17(20):2337–42

112. Ma B, Johnson R. 2012. De novo sequencing and homology searching. *Mol. Cell. Proteom.* 11(2):O111.014902

113. Welker F. 2018. Elucidation of cross-species proteomic effects in human and hominin bone proteome identification through a bioinformatics experiment. *BMC Evol. Bio.* 18(23):1–11

114. Tran NH, Zhang X, Xin L, Shan B, Li M. 2017. De novo peptide sequencing by deep learning. *PNAS* 114:8247–52

115. Clark KA, Ikram S, Evershed RP. 2016. The significance of petroleum bitumen in ancient Egyptian mummies. *Philos. Trans. A Math. Phys. Eng. Sci.* 374:20160229

116. Regert M, Bland HA, Dudd SN, van Bergen PF, Evershed RP. 1998. Free and bound fatty acid oxidation products in archaeological ceramic vessels. *Proc. R. Soc. B: Biol. Sci.* 265(1409):2027–32

117. Richnow HH, Jenisch A, Michaelis W. 1992. Structural investigations of sulphur-rich macromolecular oil fractions and a kerogen by sequential chemical degradation. *Org. Geochem.* 19(4–6):351–70

118. Colombini MP, Modugno F. 2009. *Organic Mass Spectrometry in Art and Archaeology*. Pisa, Italy: Wiley

119. Evershed RP, Heron C, Goad LJ. 1990. Analysis of organic residues of archaeological origin by high-temperature gas chromatography and gas chromatography-mass spectrometry. *Analyst* 115(10):1339

120. Regert M, Devièse T, Le Hô A-S, Rougeulle A. 2008. Reconstructing ancient Yemeni commercial routes during the Middle Ages using structural characterization of terpenoid resins. *Archaeometry* 50(4):668–95

121. Schouten S, Hopmans EC, Schefuß E, Sinninghe Damsté JS. 2002. Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? *Earth Planet. Sci. Lett.* 204(1–2):265–74

122. Matthews DE, Hayes JM. 1978. Isotope-ratio-monitoring gas chromatography-mass spectrometry. *Anal. Chem.* 50(11):1465–73

123. Dunne J, Mercuri AM, Evershed RP, Bruni S, di Lernia S. 2016. Earliest direct evidence of plant processing in prehistoric Saharan pottery. *Nat. Plants* 3:16194
124. Dudd SN, Evershed RP. 1998. Direct demonstration of milk as an element of archaeological economies. *Science* 282(5393):1478–81
125. Corr LT, Sealy JC, Horton MC, Evershed RP. 2005. A novel marine dietary indicator utilising compound-specific bone collagen amino acid $\delta^{13}C$ values of ancient humans. *J. Archaeol. Sci.* 32(3):321–30
126. Styring AK, Sealy JC, Evershed RP. 2010. Resolving the bulk $\delta^{15}N$ values of ancient human and animal bone collagen via compound-specific nitrogen isotope analysis of constituent amino acids. *Geochim. Cosmochim. Acta.* 74(1):241–51
127. Outram AK, Stear NA, Bendrey R, Olsen S, Kasparov A, et al. 2009. The earliest horse harnessing and milking. *Science* 323(5919):1332–35
128. Pääbo S. 2014. The human condition—a molecular approach. *Cell* 157(1):216–26
129. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43–49
130. Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512(7513):194–97
131. Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. 2015. Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* 16(6):359–71
132. Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. 2017. Tracing the peopling of the world through genomics. *Nature* 541(7637):302–10
133. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91
134. Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, et al. 2018. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* 553(7687):203–7
135. Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, et al. 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336(6080):466–69
136. Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, et al. 2018. The genomic history of southeastern Europe. *Nature* 555:197–203
137. Cooper A, Turney C, Hughen KA, Brook BW, McDonald HG, Bradshaw CJA. 2015. Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. *Science* 349(6248):602–6
138. Haile J, Froese DG, Macphee RDE, Roberts RG, Arnold LJ, et al. 2009. Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *PNAS* 106(52):22352–57
139. Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, et al. 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *PNAS* 104(39):15276–81
140. Frantz LAF, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, et al. 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* 352(6290):1228–31
141. Thomson VA, Lebrasseur O, Austin JJ, Hunt TL, Burney DA, et al. 2014. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. *PNAS* 111(13):4826–31
142. Gaunitz C, Fages A, Hanghøj K, Albrechtsen A, Khan N, et al. 2018. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* 360:111–14
143. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, et al. 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *PNAS* 111(52):E5661–69
144. Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, et al. 2017. Ancient genomic changes associated with domestication of the horse. *Science* 356(6336):442–45
145. Roffet-Salque M, Regert M, Evershed RP, Outram AK, Cramp LJE, et al. 2015. Widespread exploitation of the honeybee by early Neolithic farmers. *Nature* 527(7577):226–30
146. Taubenberger JK, Reid AH, Janczewski TA, Fanning TG. 2001. Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356(1416):1829–39
147. Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, et al. 2015. Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* 163(3):571–82

148. Minnikin DE, Lee OY-C, Wu HHT, Besra GS, Donoghue HD. 2012. Molecular biomarkers for ancient tuberculosis. In *Understanding Tuberculosis: Deciphering the Secret Life of the Bacilli*, ed. PJ Cardona, pp. 3–36. Rijeka, Croatia: InTech

149. Adler CJ, Dobney K, Weyrich LS, Kaidonis J, Walker AW, et al. 2013. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* 45(4):450–55, 455e1

150. Weyrich LS, Duchene S, Soubrier J, Arriola L, Llamas B, et al. 2017. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* 544(7650):357–61

151. Buckley M. 2015. Ancient collagen reveals evolutionary history of the endemic South American "ungulates." *Proc. Biol. Sci.* 282(1806):20142671

152. Westbury M, Baleka S, Barlow A, Hartmann S, Paijmans JLA, et al. 2017. A mitogenomic timetree for Darwin's enigmatic South American mammal *Macrauchenia patachonica*. *Nat. Commun.* 8:15951

153. Cleland TP, Schroeter ER, Feranec RS, Vashishth D. 2016. Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics. *Proc. Biol. Sci.* 283:20160593

154. Buckley M. 2013. A molecular phylogeny of Plesiorycteropus reassigns the extinct mammalian order "Bibymalagasia." *PLOS ONE* 8(3):e59614

155. Evershed RP, Payne S, Sherratt AG, Copley MS, Coolidge J, et al. 2008. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* 455(7212):528–31

156. Salque M, Bogucki PI, Pyzel J, Sobkowiak-Tabaka I, Grygiel R, et al. 2012. Earliest evidence for cheese making in the sixth millennium BC in northern Europe. *Nature* 493(7433):522–25

157. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528(7583):499–503

158. Copley MS, Berstan R, Dudd SN, Docherty G, Mukherjee AJ, et al. 2003. Direct chemical evidence for widespread dairying in prehistoric Britain. *PNAS* 100(4):1524–29

159. Cerling TE, Wang Y, Quade J. 1993. Expansion of C_4 ecosystems as an indicator of global ecological change in the late Miocene. *Nature* 361(6410):344–45

160. Freeman KH, Colarusso LA. 2001. Molecular and isotopic records of C_4 grassland expansion in the late Miocene. *Geochim. Cosmochim. Acta* 65(9):1439–54

161. Brocks JJ, Logan GA, Buick R, Summons RE. 1999. Archean molecular fossils and the early rise of eukaryotes. *Science* 285(5430):1033–36

162. Summons RE, Jahnke LL, Hope JM, Logan GA. 1999. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* 400(6744):554–57

163. Love GD, Grosjean E, Stalvies C, Fike DA, Grotzinger JP, et al. 2009. Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* 457(7230):718–21

164. Brassell SC, Eglinton G, Marlowe IT, Pflaumann U, Sarnthein M. 1986. Molecular stratigraphy: a new tool for climatic assessment. *Nature* 320(6058):129–33

165. Sluijs A, Schouten S, Pagani M, Woltering M, Brinkhuis H, et al. 2006. Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum. *Nature* 441(7093):610–13

166. Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, et al. 2018. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555:190–96

167. Gawad C, Koh W, Quake SR. 2016. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* 17(3):175–88

168. Schroeter ER, DeHart CJ, Cleland TP, Zheng W, Thomas PM, et al. 2017. Expansion for the *Brachylophosaurus canadensis* collagen i sequence and additional evidence of the preservation of Cretaceous protein. *J. Proteome Res.* 16(2):920–32

169. Parker GJ, Leppert T, Anex DS, Hilmer JK, Matsunami N, et al. 2016. Demonstration of protein-based human identification using the hair shaft proteome. *PLOS ONE* 11(9):e0160653

170. Zanolli C, Hourset M, Esclassan R, Mollereau C. 2017. Neanderthal and Denisova tooth protein variants in present-day humans. *PLOS ONE* 12(9):e0183802

171. Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, et al. 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409(6820):610–14

172. Vajda V, Pucetaite M, McLoughlin S, Engdahl A, Heimdal J, Uvdal P. 2017. Molecular signatures of fossil leaves provide unexpected new evidence for extinct plant relationships. *Nat. Ecol. Evol.* 1(8):1093–99

173. Hochberg GKA, Thornton JW. 2017. Reconstructing ancient proteins to understand the causes of structure and function. *Annu. Rev. Biophys.* 46:247–69

174. Randall RN, Radford CE, Roof KA, Natarajan DK, Gaucher EA. 2016. An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat. Commun.* 7:12847

175. Römpler H, Rohland N, Lalueza-Fox C, Willerslev E, Kuznetsova T, et al. 2006. Nuclear gene indicates coat-color polymorphism in mammoths. *Science* 313(5783):62

176. Pearson A. 2014. Lipidomics for geochemistry. In *Treatise on Geochemistry*, ed. H Holland, K Turekian, pp. 291–336. Oxford, UK: Elsevier

177. Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, et al. 2015. Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLOS ONE* 10(6):e0129102