Ancient DNA analysis

Ludovic Orlando[1] ✉, Robin Allaby[2], Pontus Skoglund[3], Clio Der Sarkissian[1], Philipp W. Stockhammer[4,5], María C. Ávila-Arcos[6], Qiaomei Fu[7], Johannes Krause[5], Eske Willerslev[8,9,10,11], Anne C. Stone[12] and Christina Warinner[5,13]

Abstract | Although the first ancient DNA molecules were extracted more than three decades ago, the first ancient nuclear genomes could only be characterized after high-throughput sequencing was invented. Genome-scale data have now been gathered from thousands of ancient archaeological specimens, and the number of ancient biological tissues amenable to genome sequencing is growing steadily. Ancient DNA fragments are typically ultrashort molecules and carry extensive amounts of chemical damage accumulated after death. Their extraction, manipulation and authentication require specific experimental wet-laboratory and dry-laboratory procedures before patterns of genetic variation from past individuals, populations and species can be interpreted. Ancient DNA data help to address an entire array of questions in anthropology, evolutionary biology and the environmental and archaeological sciences. The data have revealed a considerably more dynamic past than previously appreciated and have revolutionized our understanding of many major prehistoric and historic events. This Primer provides an overview of concepts and state-of-the-art methods underlying ancient DNA analysis and illustrates the diversity of resulting applications. The article also addresses some of the ethical challenges associated with the destructive analysis of irreplaceable material, emphasizes the need to fully involve archaeologists and stakeholders as part of the research design and analytical process, and discusses future perspectives.

Ancient DNA
(aDNA). Ultrashort and degraded DNA fragments that are preserved in subfossil material, including hard tissues, such as bones, teeth and shells, and soft tissues, such as mummified skin and hair, as well as sediments.

Holobiomes
The total sum of the DNA fragments making up the genome of a host organism and all of its microbiota.

✉e-mail: *ludovic.orlando@univ-tlse3.fr*

https://doi.org/10.1038/s43586-020-00011-0

In 1984, short DNA fragments were extracted and sequenced from the dried muscle of a museum specimen of the quagga, a species of zebra that became extinct at the beginning of the twentieth century, marking the birth of ancient DNA (aDNA) research[1]. Although the data gathered were limited to 229 bp, this was the first time that direct genetic information had travelled through time, adding molecular evidence to the toolkit used by researchers to observe evolution and understand how the modern world came to be.

More than three decades later, the time range amenable to DNA analysis extends to more than half a million years (560,000–780,000 years ago[2]), and many extinct species have now had their genomes completely sequenced, including woolly mammoths[3] and cave bears[4], as have human populations from Vikings[5] to Paleo-Inuit[6] to Neanderthals[7–10]. With the addition of aDNA data, our current atlas of genetic variation is not limited to a snapshot of the diversity found in present-day populations across the world. Instead, it is continuously enriched with temporal information tracking changes in the genetic ancestries of human, animal, plant and even microbial populations as they expanded, collapsed and adapted to new local environmental conditions[11–13].

aDNA has led to the discovery of new branches within the human family tree, including that of the Denisovans, who are close relatives of Neanderthals[14–16]. As a result, the genomic consequences of population decline[17–19] and the underlying environmental[20–22] and/or anthropogenic[23,24] drivers of extinctions have been revealed and clarified. Applying aDNA techniques to archaeozoological and archaeobotanical remains has also considerably enhanced our understanding of the transition from hunting and gathering to herding and farming, including how past human groups domesticated wild species showing preferred characteristics[12]. Domestication provided new opportunities for zoonotic transfer of animal pathogens to humans, which have also been analysed using ancient genomic investigation[13]. More generally, aDNA data have transformed our understanding of the historical emergence, virulence and spread of major infectious diseases[25]. Beyond the genomes of the hosts and their pathogens, the metagenomic characterization of our microbial self[26,27] and the identification of epigenetic marks[28,29] have paved the way towards a study of ancient holobiomes, which promise to reveal a deeper understanding of past social, dietary and environmental shifts and their impact on the health of individuals and populations.

Author addresses

[1]Centre d'Anthropobiologie et de Génomique de Toulouse, Université Paul Sabatier Toulouse III, Faculté de Médecine Purpan, Toulouse, France.
[2]School of Life Sciences, University of Warwick, Coventry, UK.
[3]Ancient Genomics Laboratory, Francis Crick Institute, London, UK.
[4]Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig Maximilian University, Munich, Germany.
[5]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany.
[6]International Laboratory for Human Genome Research, National Autonomous University of Mexico, Queretaro, Mexico.
[7]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing, China.
[8]Lundbeck Foundation GeoGenetics Center, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark.
[9]Welcome Trust, Sanger Institute, Hinxton, UK.
[10]The Danish Institute for Advanced Study, University of Southern Denmark, Odense, Denmark.
[11]Department of Zoology, University of Cambridge, Cambridge, UK.
[12]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA.
[13]Department of Anthropology, Harvard University, Cambridge, MA, USA.

None of these developments would have been possible without next-generation sequencing (NGS)[30], which remains thus far the most transformative technology in the history of aDNA research, profoundly affecting wet-laboratory and dry-laboratory activities alike (FIG. 1). At its most basic level, the success of NGS lies in its ability to accommodate the sequencing and analysis of millions of loci in parallel from minute amounts of ultrashort DNA fragments[31–33]. As a result, cumbersome PCR-based analyses of individual loci, which were instrumental in initially establishing the field of aDNA analysis, are now almost entirely superseded by genome-scale studies, with the exception of mini-barcodes that remain useful for characterizing past environmental communities[34]. NGS also importantly allows the study of DNA in its entirety, a cornerstone in modern authentication approaches of aDNA data. In-solution target-enrichment techniques probing specific genomic regions have become instrumental for the cost-effective recovery of sequence data at the gene[35], mitogenome[36], chromosomal[37] and even whole-genome[38] scale.

As a complement to more sensitive techniques for extracting[39–42] and integrating DNA into NGS libraries[16,43], new sources of archaeological material have broadened the range of aDNA applications. These include dental calculus[26,44], wood[45,46], mollusc shells[47] and sediments[48,49], as well as biocultural archives such as parchments[50] and textiles[51]. The discovery that particular bone types exhibit better molecular preservation[52,53] has also facilitated the collection of aDNA time-series data at both population and genome-wide scales. NGS has furthermore revolutionized the analysis of ancient sequence data, providing statistical solutions to overcome vexing problems, such as rampant contamination[54–56] and post-mortem DNA damage[57], while still enabling the detection of subtle changes in population structure by means of genetic drift, admixture and/or selection[58].

With the addition of new laboratory and computational methods developed in the past decade, aDNA analysis has now come of age. This Primer provides an up-to-date overview of the most commonly used aDNA methods and tools, as well as their limitations, and anticipates future innovations that will help propel us beyond the current state of the art as we continue to advance the scope of molecular archaeogenetics.

Experimentation

Analysing aDNA requires the destruction of irreplaceable, finite subfossil material that is part of humanity's biocultural heritage. As such, aDNA studies require careful scientific and ethical planning and a commitment to responsible research. Once a research plan is established, sample analysis proceeds similarly for most sample types. Samples are taken to a dedicated clean laboratory facility, and DNA is typically freed from its parent material by a combination of demineralization and digestion. Often, samples have high levels of degradation and are prone to contamination by other DNA sources. Following purification and concentration steps, aDNA is then constructed into an NGS library and sequenced. In addition to these core steps, there are also optional steps that can be performed at various stages to achieve desired effects, such as the complete or partial removal of post-mortem damage prior to library construction or the enrichment of specific genetic targets prior to sequencing.

Material types. Early aDNA studies considered macroscopic preservation a predictor of molecular preservation and consequently focused on soft tissues from naturally or artificially mummified remains and from stuffed or fluid-preserved museum specimens[1,59]. Hair shafts provided the DNA source for the first successfully sequenced ancient mammoth[60] and human genomes[6], in part owing to keratin's low permeability to contaminants and because it is easier to decontaminate[61]. Soft tissues, however, rarely preserve and — with the exception of hair — are usually heavily contaminated by environmental microorganisms[62]. Mineralized tissues are more abundant and typically better preserved than soft tissues, and consequently the focus of more recent aDNA studies has shifted to vertebrate bones and teeth[63]. The petrous bone has been singled out for retaining a high degree of endogenous DNA preservation[52], with ear ossicles[64] and teeth providing other suitable alternatives. Although tooth cementum can contain high amounts of host DNA[53], dentine is generally preferred for genetic analysis as, in addition to host DNA, it also allows for the recovery of ancient blood-borne pathogens[65,66].

In studies of ancient plants, suitable materials for genomic analysis include desiccated, charred, waterlogged or mineralized[67] pollen[68], cobs[69–71], pips[72], herbarium specimens[73] or seeds[74,75]. Seeds have also been found to preserve RNA[76,77]. New aDNA reservoirs are also still being discovered in the form of purely biological substrates, such as insects[78], feathers[79], eggshells[80], mollusc shells[47] or wood[45,46], but also from 'cultural' artefacts including livestock skin parchments[50] and drinking horns[81], pottery[82] or birch pitch mastics[83,84]. Beyond DNA from single focal species, whole communities can also be retrieved from single samples, such as preserved faeces

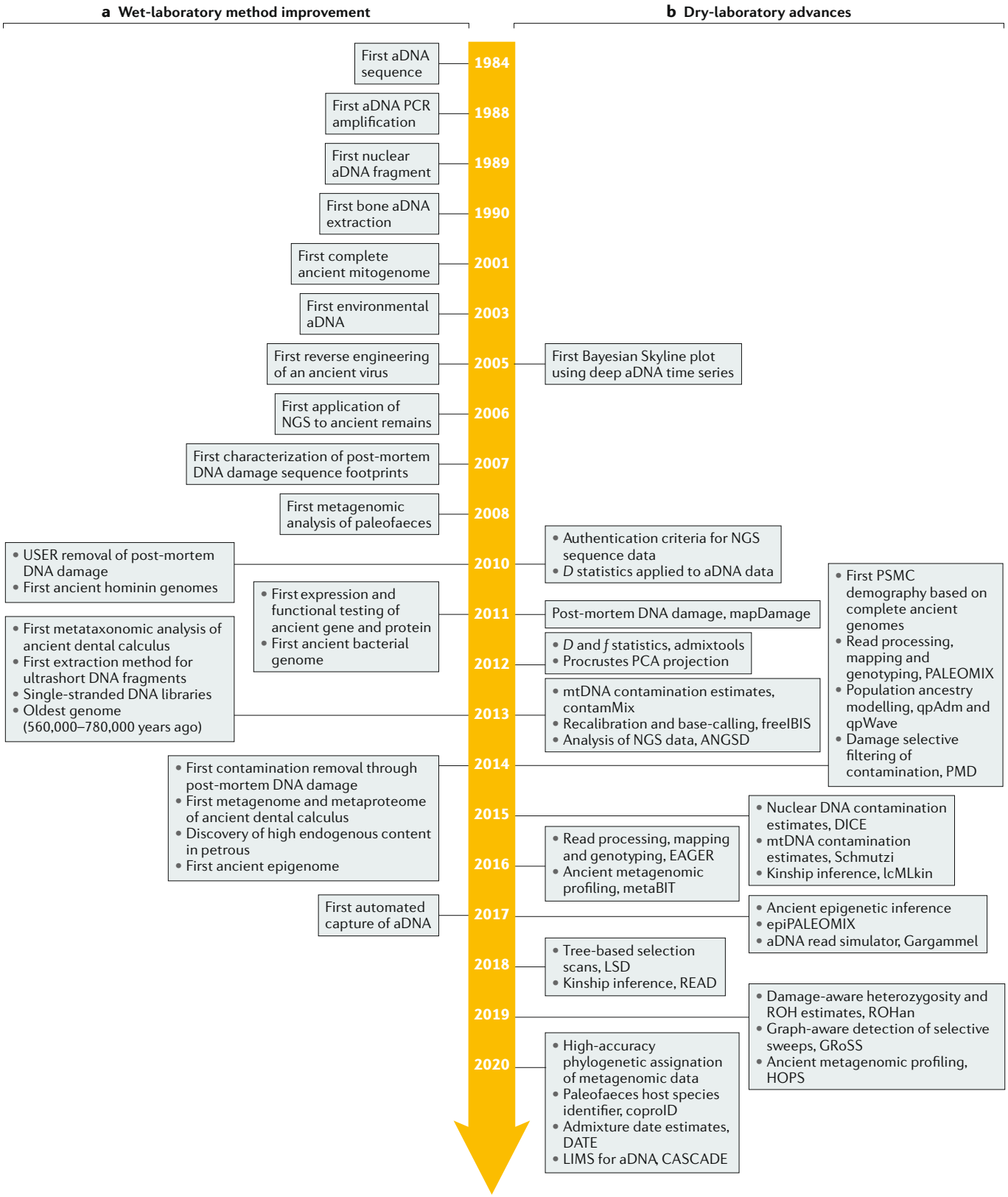


Fig. 1 | **Analytical milestones in aDNA research.** Key milestones pertaining to wet-laboratory method improvement (part **a**) or dry-laboratory advances (part **b**). aDNA, ancient DNA; LIMS, laboratory information management systems; mtDNA, mitochondrial DNA; NGS, next-generation sequencing; PCA, principal component analysis; PSMC, pairwise sequentially Markovian coalescent; ROH, runs of homozygosity; USER, uracil–DNA–glycosylase (UDG) and endonuclease VIII (Endo VIII) (New England Biolabs).

(paleofaeces or coprolites) and calcified dental plaque (calculus), allowing metagenomic analyses of the gut[85,86] and oral microbiota[26,44], respectively, as well as the detection of pathogens[26,44], parasites[87] and foods[88,89]. At larger scales, entire paleoecosystems can be reconstructed from environmental archives, such as sediments[22,49,90], ice[91] and lake cores[34,48,92].

Ethical aDNA research. There are several ethical considerations that researchers must evaluate before embarking on destructive analysis of irreplaceable archaeological material (FIG. 2). The sampling and destruction of human remains for aDNA research needs to consider cultural, historical and even political implications. Consultation with appropriate local stakeholders, including descendent communities, both prior to the start of a project and during its development, enriches aDNA research and is highly encouraged. This is not limited to human remains but applies to aDNA studies generally, including studies of animals, plants and artefacts.

Ethical issues with regard to aDNA research encompass conceptualization, sampling and communication. First, the conceptualization of the project should proceed in a way that is mindful of historical, cultural and political realities and that involves communication with local stakeholders (excavators, curators/museums, local communities and religious institutions) on equal footing. This should include detailed communication of the relevance and potential outcomes of the analyses as well as possible risks[93,94]. Second, the strategy and process of sampling should include sample documentation (for example, by photography and/or surface scanning or computed tomography[95]), the use of minimally destructive analytical techniques[96–99] and a plan for proper export and storage of samples in accordance with official permissions. The remaining sample material should be returned to the museum or appropriate group, with restoration (or production of a copy) of destroyed parts[94,100,101] when requested. When working with rare samples, the methods should first be extensively tested and demonstrated to be successful on similar but more commonly available material (for example, faunal remains at hominin sites), and sample material should not be fully exhausted but at least partially conserved for future research. Third, communication and consultation should be ongoing with stakeholders, including descendant communities. Ideally, this communication and consultation will include joint decision-making and local capacity-building, collaboration on publications and agreement about data management[102]. Wagner et al.[103] make five recommendations for establishing successful collaborations between aDNA researchers and descendant communities: consult formally with communities; address cultural and ethical considerations; engage communities and support capacity-building; develop plans to report results and manage data; and develop plans for long-term responsibility and stewardship. Finally, researchers should be mindful of the language they use in publications and specifically the potential for misinterpretation or offence given by technical terms — such as specimen, admixture or inbreeding coefficient — that may imply different meanings in scientific and general contexts. Researchers should use a reflected vocabulary that avoids biological essentialism[104] and ensure that their work avoids falling into the pitfalls of racial, nationalist or simplistic narratives[105–107] (BOX 1).

aDNA facilities. Preserved aDNA is often limited in quantity and has a high degree of degradation, which can be compounded by contamination with modern DNA. Under ideal conditions, samples would be collected immediately from excavated individuals, with minimal handling to limit potential contamination from modern sources, including staff and storage facilities. However, this is not possible for the large archives of already excavated remains in museums and institutes around the world. As such, the field of aDNA has developed various techniques to identify, remove or reduce contamination introduced during post-excavation storage and handling[40–43,54–57].

Regardless of sample origins, the extraction and manipulation of aDNA must be carried out in dedicated clean laboratory facilities to minimize further contamination risks. Such facilities are typically access-regulated and located in buildings separate from those where post-amplification DNA is manipulated. They are maintained as sterile environments through HEPA-filtered positive air pressure systems, UV exposure and daily (bleach) decontamination treatment of bench surfaces[108]. Anterooms allow researchers to dress in suitable personal protective equipment, including disposable full-body suits, gloves, sleeves, face masks and overshoes. The workspace is generally divided into multiple, separate rooms in which specific experimental tasks can be performed so as to parallelize work while limiting cross-contamination risks. Laboratory equipment is routinely decontaminated before and after use by cleaning with bleach and alcohol, whereas laminar flow hoods, with monitored air extraction and filtering systems, help prevent pollen, powder and aerosol contamination. These strict procedures are necessary to minimize modern DNA entering the facilities through reagents, ventilation and staff personnel.

DNA extraction. In order to maximize preservation of the remains' integrity and allow potential further molecular or morphological analyses, minimally destructive methods have been proposed to sample bones[96], teeth[97], insects[98] or plants[99], for example. Optimization at every experimental and computational step of the aDNA pipeline has significantly increased the sensitivity of aDNA methods, thereby decreasing the amounts of sampled material necessary for successful analyses. Small amounts of preserved DNA and extensive DNA fragmentation to lengths <100 bp[109] pose serious challenges for extracting DNA molecules from ancient samples. The reduced size of the DNA molecules makes it difficult to separate endogenous DNA of interest from contaminating DNA and from co-extracted small molecules that could act as inhibitors in downstream enzymatic reactions[110,111]. Optimized protocols have been adapted for each experimental step to accommodate various sample types and preservation states, and some allow the simultaneous

extraction of proteins for radiocarbon dating[112] and proteomic analyses[113]. In most cases, samples are first mechanically reduced to a powder and DNA is then released through incubation in buffers that decalcify mineral matrices, break down proteins and lipids, and disrupt DNA interactions with organic and inorganic compounds. Removal of loosely bound contaminants can be achieved through pre-digestion cleaning steps

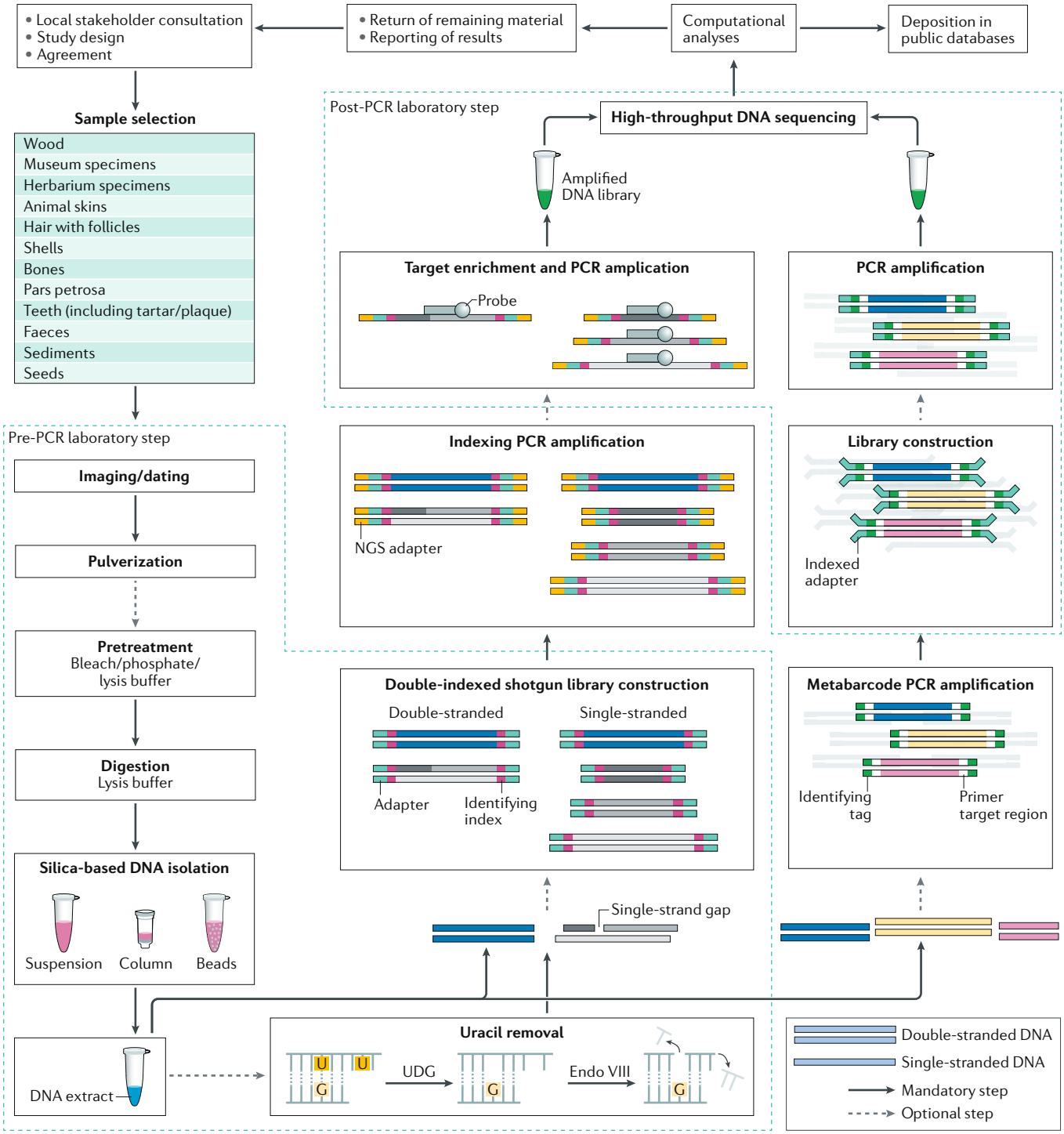


Fig. 2 | **Experimental workflow.** A wide range of remains are amenable to ancient DNA (aDNA) analysis. Prior to sample destruction, a research plan should be agreed amongst the different stakeholders. The different wet-laboratory procedures must be carried out in specific aDNA facilities, minimizing environmental contamination, and include all pre-amplification experimental steps, including sample preparation, DNA extraction, optional USER treatment and DNA library construction. Target enrichment and PCR amplification are carried out in regular molecular genetics facilities. Following next-generation sequencing (NGS), the sequence data are processed on computational servers and uploaded to public repositories. Results should be communicated to the stakeholders and any remaining sample should be returned as per the initial agreement. USER, uracil–DNA–glycosylase (UDG) and endonuclease VIII (Endo VIII) (New England Biolabs).

Box 1 | **Writing archaeogenetic prehistory**

In the past decade, archaeogenetic studies have provided radically new insights into prehistory worldwide. However, several early publications did not fully appreciate the related history of research in archaeology. The inherent complexity of relevant terms such as culture, migration or people in the social sciences was insufficiently acknowledged, in part owing to different disciplinary publishing norms and to the strict limitations on word count and number of references imposed by high-profile journals, which resulted in very negative feedback by the archaeological community. Now, sustainable collaborations between archaeogeneticists, archaeologists and historians have been established and even institutionalized (such as with the Max Planck Harvard Research Center for the Archaeoscience of the Ancient Mediterranean), and awareness has been raised on both sides — for the use of challenging terms, on the one hand, and the potential of archaeogenetic historiography on the other. Increasingly, a new generation of scholars is being trained within an interdisciplinary framework that allows narratives of the past to be told in more integrative, nuanced and sophisticated ways that address the complexity of archaeological and archaeoscientific data sets.

consisting of either a bleach treatment[40] to oxidize and destroy contaminant DNA and chemical inhibitors or a staged digestion[41,53], whereby the first extracted fraction (which is likely to contain the majority of contaminants and inhibitors) is set aside or discarded. The methods may also be combined[42], but both approaches also result in varying degrees of endogenous aDNA loss, and therefore aggressive cleaning should be viewed as a trade-off that is subject to diminishing returns.

Because it is highly fragmented, aDNA behaves differently to high molecular weight genomic DNA from fresh specimens, and requires custom protocols for efficient recovery[39]. Methods most often used for aDNA isolation rely on the adsorption of DNA molecules to silica particles[110] in a chaotropic binding buffer carried out in solution[114], on spin columns[41,42,53,115] or on suspended silica-coated magnetic beads[115,116]. DNA is then eluted in a low-salt buffer after ethanol washes. In some instances, a phenol–chloroform purification step can advantageously be used as an alternative[117] or complementary[118] approach. Optimizing DNA extraction and library construction to recover fragments as short as 35 bp has been decisive in characterizing DNA from samples showing extensive molecular degradation[39,119].

Post-mortem DNA damage removal. A wealth of chemical reactions are known to affect DNA after death and include the fragmentation of DNA molecules into ultra-short DNA fragments, the conversion of the four nucleotides into various derivatives and the cross-linking of DNA to other molecules. These post-mortem alterations potentially hinder the manipulation of aDNA and reduce the amount of retrievable genetic information[109,120]. The most dramatic of these changes is DNA fragmentation, which occurs through hydrolytic depurination followed by β-elimination reactions that 'nick' (break) the DNA backbone (FIG. 3); during this process, the DNA double helix breaks apart into pieces that can be millions of times shorter than their original length during life. The next most common hydrolytic degradation reaction is the deamination of cytosines, which are converted into uracils and thereafter sequenced as thymine analogues[121] (FIG. 3a). This process is responsible for the sequencing artefacts observed as C to T misincorporations (and also G to A misincorporations for double-stranded DNA libraries) (FIG. 3b). The rate of such misincorporations increases towards the ends of reads when mapped against a reference sequence owing to cytosine deamination preferentially occurring in the single-stranded overhanging termini of aDNA fragments[121]. Post-mortem damage can accumulate with the age of the sample[122], but degradation kinetics are ultimately driven by local environmental conditions[2,45]. As a result, some recent samples can appear more damaged than older ones. In addition, degradation processes are best studied in mineralized tissues and are less well understood in other kinds of remains.

Although C to T misincorporations provide handy genetic signatures that can be used to authenticate aDNA sequences[54,57], they can also bias sequence analyses, potentially leading to incorrect conclusions[123,124]. To lower the impact of damage-induced sequence errors, DNA extracts can be optionally treated before library construction with a commercialized enzymatic mix of uracil–DNA–glycosylase (UDG) and endonuclease VIII (Endo VIII) known as the USER reagent (New England Biolabs). This reagent removes uracils and cleaves the resulting abasic sites, thereby cutting out damage but also shortening the length of the DNA molecule[125]. Although having the beneficial effect of reducing sequencing errors, USER treatment also has the negative effect of eliminating the damage patterns that are needed to authenticate aDNA sequences and differentiate them from contaminating DNA; however, in the case of mammalian nuclear DNA, this damage signal can still be observed by examining CpG dinucleotides (FIG. 3c). For non-mammalian DNA, some laboratories first perform screening of a non-treated (non-UDG) DNA library to determine sequence authenticity, and then build a second USER-treated (full-UDG) library for analysis. However, the need to build multiple libraries per sample can dramatically increase costs. As a solution, a modified USER protocol has been developed (UDG-half) that removes most damage but retains a single uracil at each end of the molecule[126], thereby allowing aDNA authentication but limiting damage to the terminal base where it can be easily masked or clipped computationally during downstream analysis. Decisions on the use of USER treatment are generally study dependent and involve calculations regarding the relative importance of sequence fidelity, sequence length and cost for a given study design and research question.

Mini-barcode PCR amplification. One of the few areas of aDNA research that still directly amplifies raw DNA extracts by PCR is the field of palaeoenvironmental reconstruction, which attempts to recover information on past species' communities through DNA barcoding[127]. This is achieved through the use of PCR capable of amplifying target loci from a wide range of organisms, typically targeting the *trnL* loop of the chloroplast genome in plants[128], and the mitochondrial targets *cytb*[90], *CO1* (REF.[91]), *12S* or *16S*[129] in animals. Such targets are selected on the basis of their relative high variability such that each species is expected to be defined by distinct haplotypes — like barcodes — thereby distinguishing them from other species.

DNA library
A molecular construction in which DNA fragments are ligated to DNA adapters of known sequences in order to be amplified and optionally captured prior to sequencing; different sequencing platforms require different library constructs.

DNA barcoding
The taxonomic assignment of metagenomic DNA content on the basis of DNA fragments that show limited intra-specific sequence diversity but large inter-specific sequence diversity.

Shotgun sequencing
Non-targeted sequencing of DNA library content.

High taxonomic resolution is often achieved through meta-barcoding, an approach that combines barcode PCR with NGS[22,130–132] and uses tagged (indexed) primers, eliminating the need for conventional molecular cloning to separate out individual sequences from the amplicon pool. Multiple samples can be simultaneously sequenced owing to the incorporation of short sample-specific index sequences into the PCR primers, usually of 4–6 bases, thereby allowing later bioinformatic recovery of specific sample sets. This approach is popular because of its cost-effectiveness and perceived sensitivity. However, it also has limitations. First, the termini of the aDNA fragments are generally not amplified using this method, so damage parameters such as cytosine deamination cannot be used as an authentication criterion. Second, allelic drop out, the failure of certain templates to amplify, can occur as a result of the clonal biases inherent in PCR or from length polymorphisms in the target region[133], both of which are especially problematic when DNA is degraded and the template copy number is low, giving rise to low or skewed sequence complexity. As with PCR-based aDNA approaches used in the 1990s, the authentication of meta-barcoding results relies instead on reproducibility across multiple PCR replicates[128] and demonstrating that negative controls are either blank or contain different sequence signatures from that seen within the samples. Alternative shotgun sequencing and/or target-enrichment approaches are increasingly being applied in palaeoenvironmental reconstructions[48,49,134,135], along with authentication approaches suitable for low-representation metagenomic data[136]

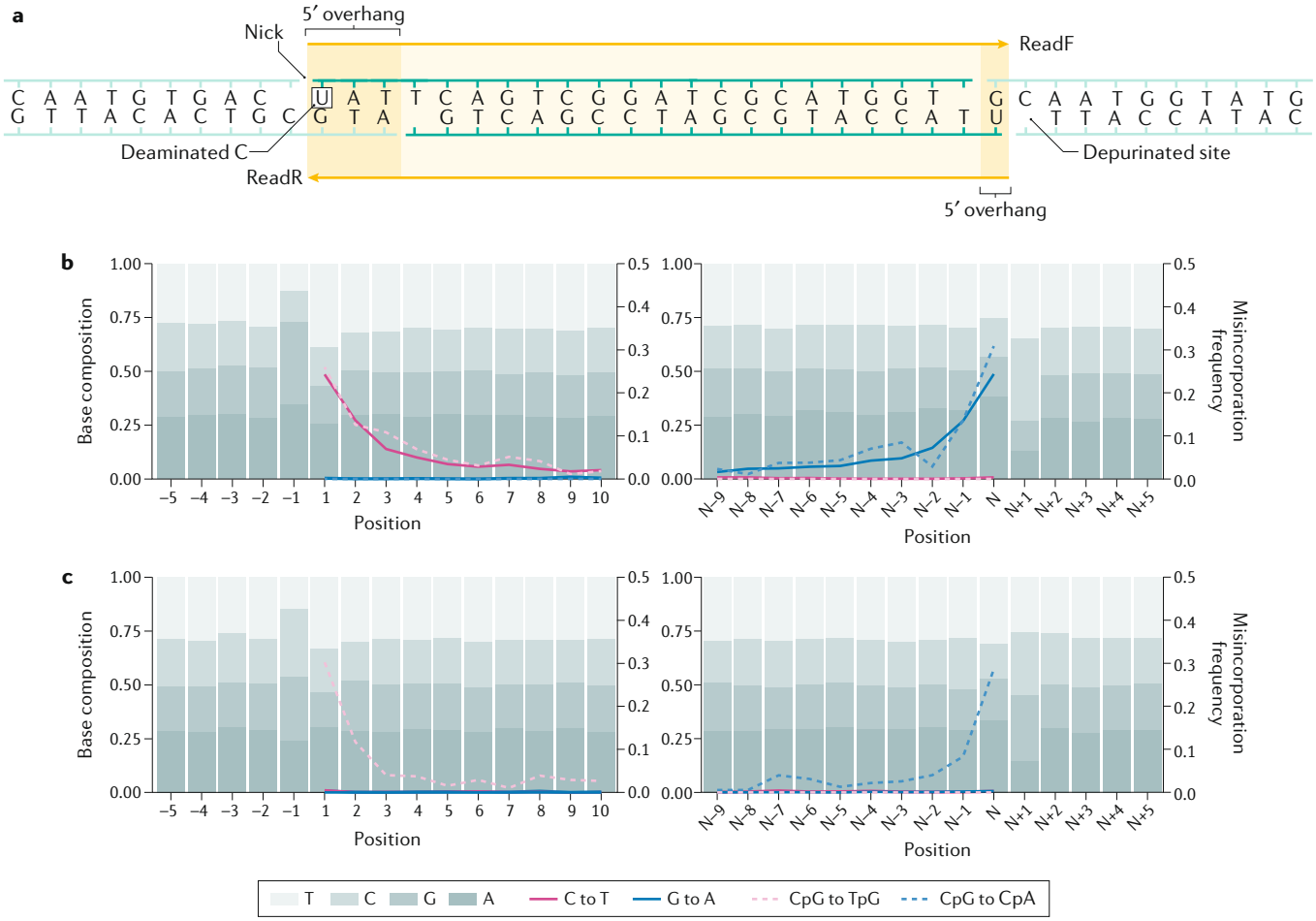


Fig. 3 | **Post-mortem DNA damage and data authentication. a** | A typical ancient DNA (aDNA) molecule. Fragmentation through purine loss results in the formation of overhanging ends, shown as 5′ overhangs, where cytosine deamination is considerably faster than within double-stranded parts. The base composition and nucleotide misincorporation profiles around read starts and read ends are shown for data generated on double-stranded DNA libraries both in the absence (part **b**) and in the presence (part **c**) of USER treatment[125,126] (data from[255]). Positions 1–10 refer to the first 10 read positions, whereas positions –5 to –1 correspond to the 5 genomic positions located upstream of reads. The base composition indicated corresponds to that of the reference genome. Positions N–9 to N refer to the last 10 read positions, whereas positions N+1 to N+5 correspond to the 5 genomic positions located downstream of reads. In the absence of USER treatment, sequence data are enriched in purines (G and A) at genomic positions flanking read starts and show an increasing excess of C to T misincorporations towards read starts, regardless of CpG contexts. The same happens with G to A at read ends. Such cytosine deamination profiles are lost for sequences generated using the same sample and methodology but applying USER to DNA extracts, except within CpG contexts. DNA processing during the construction of single-stranded DNA libraries results in different base compositional profiles, in which profiled read ends are symmetrical to those observed at read starts, instead of being reverse complemented. ReadF and ReadR form the read pair resulting from paired-end sequencing. USER, uracil–DNA–glycosylase (UDG) and endonuclease VIII (Endo VIII) (New England Biolabs).

and the use of phylogenetic assignment algorithms that are robust to poor database representations of taxa[137].

DNA library construction. Molecular cloning of PCR amplicons into bacterial plasmids was the first methodology used to build aDNA fragments into DNA libraries, which were then cultured to retrieve sufficient material for sequencing using capillary electrophoresis[138]. The limitations of capillary instruments, however, never provided sufficient sequencing throughput to obtain sequence data beyond the megabase scale[138,139]. Culture-free NGS DNA libraries provide a substantially less labour-intensive solution as they can be virtually immortalized by PCR for unlimited reuse prior to sequencing. There are two main library preparation methods currently available for sequencing by synthesis (for example, Illumina sequencing[140]), the main type of NGS used for aDNA: double-stranded library preparation and single-stranded library preparation. Each has different characteristics in terms of costs, hands-on time and sensitivity. In the double-stranded preparation method, aDNA molecules are end-repaired and ligated to double-stranded adapters[121,141,142], whereas in the single-stranded preparation method, heat-denatured aDNA templates and adapters are ligated as single-stranded molecules[16,143,144] (FIG. 2). The latter approach shows increased sensitivity for low biomass samples as overhanging 3′-OH termini and even strands with nicks are amenable to ligation. These features are enzymatically removed or improperly built using double-stranded library preparation. Adding biotinylated adapters one at a time during a single-stranded preparation also prevents the formation of constructs refractory to sequencing and reduces material loss during library purification. Another advantage of the single-stranded preparation method is that it enables the molecular selection of DNA templates carrying evidence of post-mortem DNA damage, which can increase the fraction of endogenous DNA incorporated into sequencing libraries and thus reduce downstream sequencing costs[43]. Although initially cost-prohibitive, the single-stranded methodology has now been modified to allow ligation using inexpensive double-stranded DNA ligases[145,146]. Experimental procedures in which libraries are prepared within single tubes are also available[147], which further reduces manipulation and hands-on time while facilitating parallelization[144].

All NGS aDNA libraries include short unique identifying sequences (or indexes) integrated within their adapters to ensure traceability from preparation to sequence production. These indexes allow pooling of multiple samples within sequencing runs and also greatly reduce contamination risks within the laboratory as only sequences deriving from samples should carry the appropriate adapter–index combination. Their sequence can be read as part of the main sequencing reads[126] or through external sequencing primers[142], or both[148]. Using indexes on both adapters is highly recommended to allow the detection of chimeric DNA templates that can form through jumping PCR[142] and index hopping (or index switching) during cluster generation[149].

DNA library amplification. Although single-molecule sequencing (for example, Helicos, Pacific Biosciences and Oxford Nanopore) has provided a PCR-free alternative for sequencing aDNA libraries[2], the generation of almost all aDNA data to date has required library PCR amplification to boost signals prior to sequencing. PCR, however, does not simply restore sufficient amounts of material for downstream manipulation; it can also significantly alter the library composition as DNA polymerases show various preferences for templates of particular sizes and/or base composition[150], as well as differing in 'proofreading' capabilities for bypassing post-mortem DNA damage[48,151]. Pfu Turbo Cx, Herculase II and Accuprime Pfx represent some of the most commonly used DNA polymerases for their capacity to perform even amplification of fragments with varying base composition and length, and to preserve the original library complexity. Pfu Turbo Cx is a non-proofreading enzyme capable of amplifying damaged, uracil-containing templates without enzymatic stalling and is typically used during library construction. Herculase II and Accuprime Pfx are proofreading, high-fidelity enzymes that are typically used for subsequent amplification steps during target enrichment and sequencing preparation but may also be used to produce libraries prepared from USER-treated templates. Increasing the number of PCR cycles drives the formation of PCR duplicates, resulting in clonality, and must be mitigated to avoid saturation during sequencing unless target enrichment is performed. Real-time PCR quantification of library concentration can help identify the optimal number of cycles required for final library amplification, as well as estimate the library complexity in order to decide on sequencing depth[152]. Performing independent amplifications of library aliquots in parallel also helps to maintain complexity and reduce sequencing redundancy, and further PCR reconditioning can be helpful in reducing heteroduplex formation[153]. Technical guidance for determining optimal amplification conditions, including numbers of PCR cycles, has been released for numerous protocols (for example, see REFS[143,154]).

Target enrichment. The often-limited endogenous fraction (for example, DNA of the species of interest) present in aDNA libraries can make shotgun sequencing of entire ancient genomes uneconomical, especially at high depths of coverage. Target-enrichment approaches have thus been developed to focus sequencing efforts on library content of particular interest. Early approaches relied on library annealing to microarray-bound probes[155] but now various in-solution hybridization approaches involving short DNA[35–37] or RNA[156,157] oligonucleotides are favoured. These generally consist of 52-mers[158], 60-mers[159] and 80-mers[72] and can target from thousands to millions of loci[160], including entire chromosomes[37,161] and whole genomes[38,157]. However, probe design must attend to base compositional and structural considerations[159] to avoid biasing the recovery of sequence variants. In addition, ascertainment bias introduced by the selection of particular probe sets can also affect estimates of population affinities. For example,

DNA ligases
A class of enzymes that are capable of stitching together different DNA fragments.

Ascertainment bias
Statistical bias resulting from the collection of genetic data at a subset of loci that do not reflect the overall genetic diversity present at the whole-genome scale.

targeting single-nucleotide polymorphism (SNP) variants that are common in a given population can result in overlooking non-targeted SNPs that may be more frequent in other populations. Probe ascertainment should therefore be controlled or corrected for whenever intended for population inference[162]. This can be achieved in silico for species in which extensive genome sequence data have been generated, by comparing population genetics statistics based on whole-genome data with those estimated when conditioning on specific genome locations, as well as through various modelling approaches[163].

Target-enrichment probes can be purchased directly from provider companies, or probes can also be generated in-house using oligonucleotide synthesizers or by following laboratory procedures converting fresh DNA extracts or RNA transcripts of closely related taxa into DNA libraries that can be amplified and/or transcribed to produce probes on demand[36,38,164,165]. A wide range of predesigned and custom in-solution target-enrichment kits are available through companies such as Arbor Biosciences and Agilent Technologies, but large probe sets, such as those used in comparative studies with worldwide human genotype panels[160], are commercially restricted to some laboratories. Although large probe sets are available on commercial microarrays, such as the Axiom Genome-Wide Human Origins 1 Array, microarrays require high-quality genomic DNA for input and are generally incompatible with aDNA. Although some laboratories have negotiated access to large custom in-solution probe sets on an ad hoc basis, the field as a whole would benefit from collective bargaining to make the manufacturing of large high-demand probe sets more accessible.

Performing two successive rounds of enrichment generally increases on-target coverage rates[166], whereas the use of short adapters can favour on-target library to probe annealing and enhance enrichment efficacy[126]. Further rounds of enrichment are generally not beneficial, as they tend to increase library clonality and therefore lead to diminishing returns. Finally, probe-free enrichment approaches leveraging the affinity of methyl-binding domains for methylated CpGs have been proposed to separate the microbial fraction from the vertebrate and plant DNA templates present in aDNA extracts, but their efficacy is generally limited by the short size of aDNA fragments[62,167].

DNA sequencing. Different NGS instruments have been used to retrieve aDNA data from subfossil and museum specimens. The Roche 454 system was first used in 2006 (REF.[31]) and remained dominant until 2010 (REF.[7]) when the first ancient human genome was sequenced using Illumina technology's sequencing by synthesis technology[6]. Although other technologies have been sporadically used (for example, SOLiD[168], Helicos[2], BGI-SEQ500 (REF.[169]) and Ion Torrent[170]), the vast majority of aDNA sequence data have been produced using Illumina instruments, owing to their ubiquity, high data output, cost-effectiveness and relatively low error rates. They also work optimally on relatively short DNA sequences (<300 bp) and are well suited to sequence DNA in the range of 50–150 bp, which characterizes most aDNA. Illumina platforms are not devoid of problems, however, and the assessment of possible batch effects may be required. This is usually achieved through a recalibration of individual base quality scores using spiked-in PhiX control libraries[171,172] to account for variability specific to each individual sequencing run. During sequencing preparation, it is also important to take steps to mitigate index hopping when using Illumina platforms, especially those using patterned flow cells and isothermal cluster generation by exclusion amplification, such as the HiSeq4000, HiSeqX and NovaSeq platforms[173]. These steps include removing heteroduplexes and free adapters from sequencing pools prior to sequencing and computationally removing chimeric sequences after sequencing based on dual index combination analysis[142,149].

Information management systems. Truly large-scale paleogenomics projects are now a reality for aDNA laboratories. Such projects may involve an extensive network of collaborators with large numbers of samples that are analysed using a wide range of methods. Documentation of metadata and experimental steps is thus becoming increasingly important in order to optimize work coordination and collaboration, as well as to ensure quality control and ethical handling of the samples. This information is generally digitized in most, if not all, aDNA laboratories, and laboratory information management systems (LIMS) specifically developed for aDNA, such as CASCADE[174], are becoming freely available.

Results

Numerous tools and pipelines are available for aDNA sequence analysis, and the selection of appropriate procedures depends on the nature of the research project objectives — for example, whether aimed at paleoenvironmental reconstruction, microbial profiling or modelling population history. There are, however, numerous steps in the analytical workflow that are common to most projects. These involve the processing of raw sequencing data, alignment against reference genomes and/or sequence databases, and the assessment of authenticity and error rates, including miscoding lesions resulting from post-mortem damage. Analytical tools used in aDNA analysis are presented in TABLE 1.

Read processing and alignment. DNA molecules can be sequenced in one or both directions, resulting in single reads or paired-end reads. The first processing steps include index demultiplexing to identify those reads belonging to each individual library based on their index combination, followed by read trimming to remove adapters and/or low-quality terminal sequences and, finally, the collapsing of overlapping read pairs into consensus sequences[175,176]. Although with exceptions[177], the limited length of aDNA and the presence of often dominant environmental microbial DNA contamination have precluded widespread usage of de novo sequence assembly. Processed reads are instead typically aligned using BWA[178] or Bowtie2 (REF.[179]) to reference genomes of the focal species, such as the human genome but also potential microbial pathogens that may have infected

Demultiplexing
A process by which pools of sequences originating from different DNA libraries are assigned back to their original samples on the basis of short synthetic sequences added during library indexing.

Table 1 | **Analytical tools used in aDNA analysis**

Software	Link	Description
CASCADE[174]	Available by contacting authors	LIMS for aDNA experimental workflow
PALEOMIX[187]	https://paleomix.readthedocs.io/en/latest/	Read alignment and processing, phylogenomics
nf-core/ EAGER[188]	https://eager.readthedocs.io/en/latest/index.html https://github.com/nf-core/eager	Read alignment and processing
mapDamage2 (REF.[57])	https://ginolhac.github.io/mapDamage	Post-mortem DNA damage assessment
PMDtools[55]	https://github.com/pontussk/PMDtools	Selection of reads showing signatures of post-mortem DNA damage
Schmutzi[189]	https://grenaud.github.io/schmutzi/	Contamination estimates based on mitochondrial DNA data
DICE[346]	https://github.com/grenaud/dice	Contamination estimates based on nuclear data
VerifyBamID[347]	https://github.com/statgen/verifyBamID/releases	Identification of contamination and/or sample swaps
Gargammel[180]	https://grenaud.github.io/gargammel/	aDNA read simulator
metaBIT[242]	https://bitbucket.org/Glouvel/metabit/src/master/	Taxonomic profiling of (ancient) metagenomic data
HOPS[241]	https://github.com/rhuebler/HOPS	Taxonomic profiling of (ancient) metagenomic data
MEx-IPA	https://github.com/jfy133/MEx-IPA	Interactive viewer of MALT taxonomic assignments
coproID[233]	https://github.com/nf-core/coproid	Identification of the host sources of faecal material
epiPALEOMIX[247]	https://bitbucket.org/khanghoj/epipaleomix/wiki/Home	Inference of aDNA methylation and nucleosome mapping
DamMet[248]	https://github.com/KHanghoj/DamMet	Inference of aDNA methylation, accounting for DNA damage, sequencing and genotyping errors
ANGSD[191]	http://www.popgen.dk/angsd/index.php/ANGSD	Variant identification, population genetics inference
ATLAS[348]	https://bitbucket.org/wegmannlab/atlas/wiki/Home	Variant identification
ADMIXtools[219]	https://github.com/DReichLab/AdmixTools	Population genetics inference
smartPCA	https://github.com/chrchang/eigensoft/wiki/smartpca	PCA and Procrustes PCA projection
bammds[213]	https://savannah.nongnu.org/projects/bammds/	Multidimensional scaling
PCAngsd[209]	http://www.popgen.dk/software/index.php/PCAngsd	PCA, admixture and selection signatures
DATES[253]	https://github.com/priyamoorjani/DATES	Inference of admixture timing
LSD[349]	https://bitbucket.org/plibrado/LSD/src	Selection signatures
GRoSS[350]	https://github.com/FerRacimo/GRoSS	Selection signatures
ROHan[351]	http://grenaud.github.io/ROHan/	Heterozygosity estimates and runs of homozygosity
hapROH[228]	https://pypi.org/project/hapROH/	Inbreeding inference from low-coverage data
lcMLkin[201]	https://github.com/COMBINE-lab/maximum-likelihood-relatedness-estimation	Kinship inference
READ[203]	https://bitbucket.org/tguenther/read/src/master/	Kinship inference
SourceTracker[243]	https://github.com/danknights/sourcetracker	Metagenomic authentication

aDNA, ancient DNA; LIMS, laboratory information management systems; PCA, principal component analysis.

the individuals during life. Parameters maximizing mapping specificity and sensitivity can be identified using aDNA read simulators such as gargammel[180], and depend on the evolutionary distance to the reference genome used, the DNA fragment size and the amount of post-mortem damage[111,181–184]. Mappers can show different performance depending on the experimental procedure followed during sample preparation, such as the type of DNA library constructed[185] and whether or not extracts were USER-treated[184].

Single haploid genomes provide only one version of the structural and sequence diversity present in a species. Mapping to single reference genomes can thus introduce substantial reference bias, especially as short

reads carrying the alternate allele can prove particularly difficult to map[162]. Mitigation solutions include the use of variation-aware read aligners such as vg[186] or read filtering procedures post mapping[148]. Bioinformatic pipelines such as Paleomix[187] and EAGER[188] provide automated solutions to convert raw sequencing reads into alignments available for downstream analyses, and more.

Authentication. Authentication of genomic DNA from single individuals (see below for authentication of microbial DNA) requires both the observation of post-mortem nucleotide misincorporation arising from cytosine deamination[57,120] (FIG. 3b,c) and the estimation of contaminating DNA based on heterozygosity levels found in haploid chromosomes[54]. Post-mortem DNA damage patterns can be quantified in non-USER-treated or partial USER-treated libraries using the tool mapDamage2 (REF.[57]). Cytosine deamination can also be estimated using the tool PMDtools[55] on data generated following full USER treatment by conditioning on CpG dinucleotides[125], as these are heavily methylated in the nuclear genomes of vertebrates and protected from USER degradation. Estimating the degree of contamination using ploidy can be challenging as it generally requires access to a polymorphism database to account for the probability that contaminating molecules are undetected because the contaminant and the endogenous genome share the same variants. However, multiple tools exist for assessing contamination levels in haploid chromosomes, both from mitochondrial DNA data (for example, schmutzi[189]) and using the X chromosome in males (for example, ANGSD[190,191]). Emerging methods such as AuthentiCT[192], which uses a model of cytosine deamination to estimate contamination, promise to generalize contamination inference across organisms, sex and chromosome types, thereby mitigating the need for the ploidy-based approaches. Additional approaches for estimating contamination work by isolating sequences with evidence of cytosine deamination[55,193,194] and then detecting differences in ancestry[194] or linkage disequilibrium[195] between the damaged (alleged ancient) fraction and the total data using reference panels.

Error rates. The total number of nucleotide misincorporations present in a reconstructed genome depends on the amount of sequencing data produced, the amount of post-mortem DNA damage that has accumulated and whether or not USER treatment was used in the experimental workflow. Reconstructed ancient genomes generally contain sequences with both genuine and artefactual variants, which may impact population-level analyses of allelic sharedness and genetic distances. Relative error rates can be estimated by comparing the genetic distance of an ancient specimen with an outgroup: the inflation of genetic distance between the ancient genome and the outgroup genome relative to the genetic distance separating the same outgroup and a high-quality modern genome provides a conservative measure of the error rate[2,190]. Breaking down individual error rates per nucleotide substitution type (for example, C to T, G to A, C to A and G to T) generally reveals a pervasive excess of transitions (C to T, G to A and their reciprocal changes) owing to post-mortem cytosine deamination[7]. This can be done within the framework of ANGSD software[191] that covers a wide analytical range specific for NGS and aDNA data, including the calculation of genotype likelihood scores that are aware of individual base qualities (such as the confidence in each base called during sequencing) and many of the analyses presented below. Read simulation with gargammel[180], conditioning calculations on transversions[7] or other substitution classes showing non-inflated error rates[148], filtering CpGs[196] (FIG. 3b,c) or subtracting total counts of mutational patterns proportionally to error rates[197] can help ensure the validity of downstream conclusions.

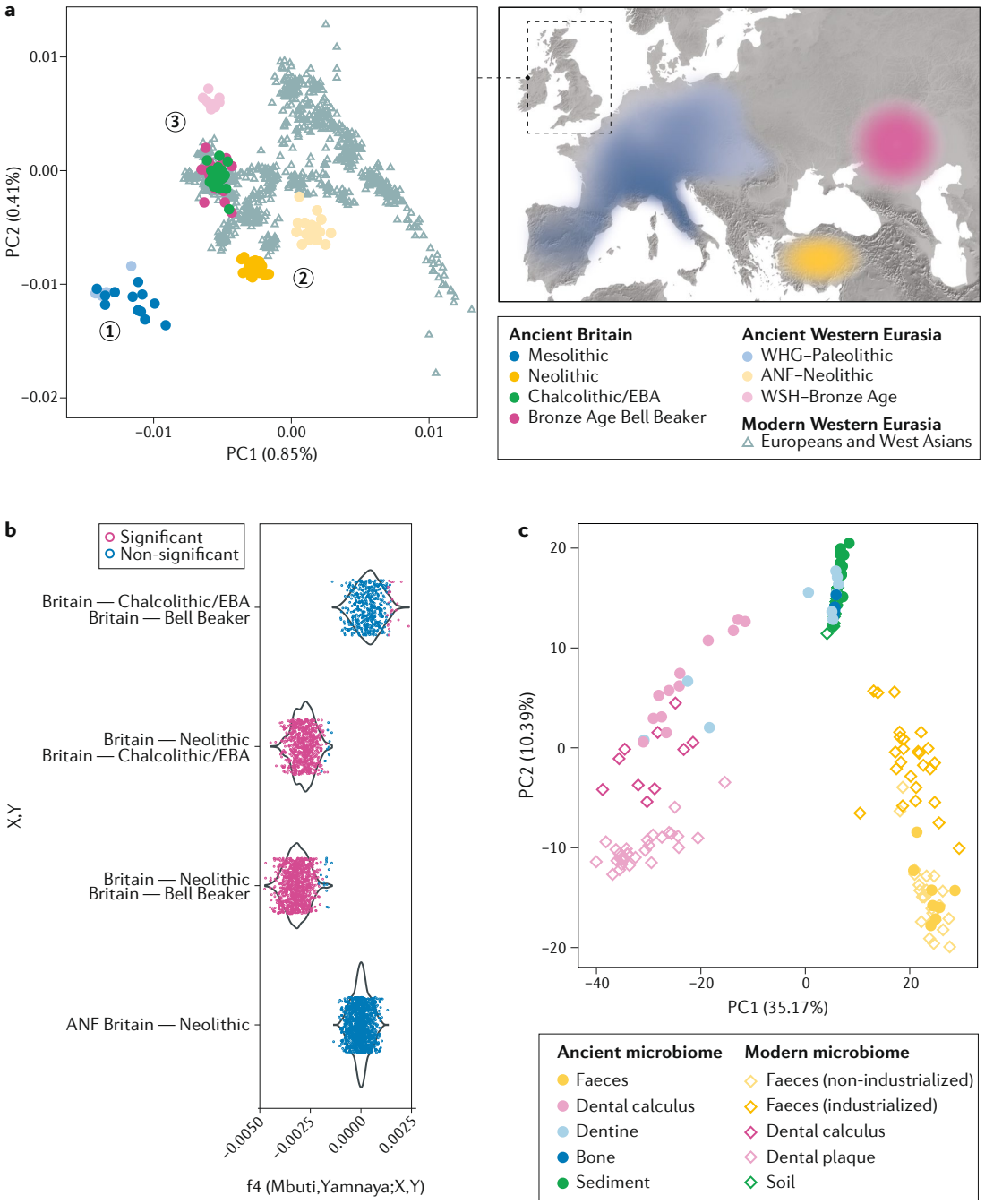
Molecular sexing and kinship. Biological sex, as determined by the karyotype of sex chromosomes in species such as humans (XY in males, XX in females and exceptional karyotypes such as XXY, XYY and others), is straightforward to estimate even with only a few thousand random sequences from a human individual[198]. To account for misalignments between the sex chromosomes, different approaches compare the number of sequences aligning to the X, Y and autosomal chromosomes in different ways to establish calibrated expected values for each biological sex[16,198,199]. For species in which reference assemblies lack the Y chromosome, X to autosomal chromosome coverage ratios are often used[200], and require no more than 1,000–10,000 high-quality read alignments.

Methods detecting genetic kinship generally make use of population panels in which expectations of identity by descent (or allelic mismatch rates) can be derived in unrelated individuals[201–203]. They are implemented in user-friendly tools, such as lcMLkin[201] and READ[203], providing inferences up to second-degree relationships and potentially more. The latter, however, requires that at least three individuals from a given group are available to assess the sequence similarity thresholds corresponding to similar kin. Together with observed shared matrilineal and patrilineal markers on the mitochondrial DNA and Y chromosomes (which require a minimal 3–5× sequence coverage[204]), extensive pedigrees can then be reconstructed[205]. However, such approaches are sensitive to the allelic frequencies within the population panel used, which ideally should be close to the allelic frequencies of the population investigated. The need for a population baseline can be avoided by computing individual heterozygosity at loci with coverage >1× (REF.[206]).

Multivariate analyses. Principal component analysis (PCA) is a classic, exploratory statistical method that is used to represent genetic affinities among individuals within simple graphs. It summarizes genetic variation measured from individual genotypes in hundreds to thousands of individuals for thousands to millions of SNP loci into a reduced number of dimensions that are shaped by ancestry[207,208]. Those dimensions represent the principal components and provide the main axes of the graphical representation of genetic affinities. aDNA data from an individual are often projected onto present-day genetic variation[194] (FIG. 4a), but when sufficient data are available, ancient specimens can be

Outgroup
An individual, a population or a group of populations and/or species that are genetically close but different from those under study.

Identity by descent
DNA segments between two or more individuals are identical by descent when they are inherited from a common ancestor in the absence of recombination.



included in the principal component computation itself. In cases where ancient genetic coverage is low, it is generally necessary to pseudo-haploidize modern data by randomly selecting one read at each polymorphic position, unless approaches are used that explicitly account for varying coverage across sites and individuals[209]. Differences in coverage across ancient individuals can also be mitigated using Procrustes analysis of independent PCAs[194] followed by least-squares projection[210]. It is notable that PCA clustering is sensitive to the sample size of different ancestries and also to their amount of genetic drift, which exaggerates principal component distances[207,208,211,212]. Therefore, individuals from the same population but distant in time may be misleadingly separated in PCA space. Multidimensional scaling has been proposed as an alternative to PCA in cases where only minimal sequence coverage is available, typically 0.001–0.1× (REF.[213]) for human data. For both methods, accurate clustering can be achieved even in the absence of sequence overlap between different ancient individuals, as long as sufficient sequence data enable the estimation of genetic distance to a predefined panel of relevant ancestries.

Population ancestry modelling. Whereas PCA results are most often visualized as biplots and in practice analysed in 2D comparisons of two principal components at a time, alternative clustering methods such as those implemented in ADMIXTURE[214] make higher-dimensional assessments possible. There is a broad suite of software

Procrustes analysis
Also known as Procrustes superimposition. A statistical method allowing the translation, rotation and scaling of multidimensional objects within a single analytical space where they can be compared.

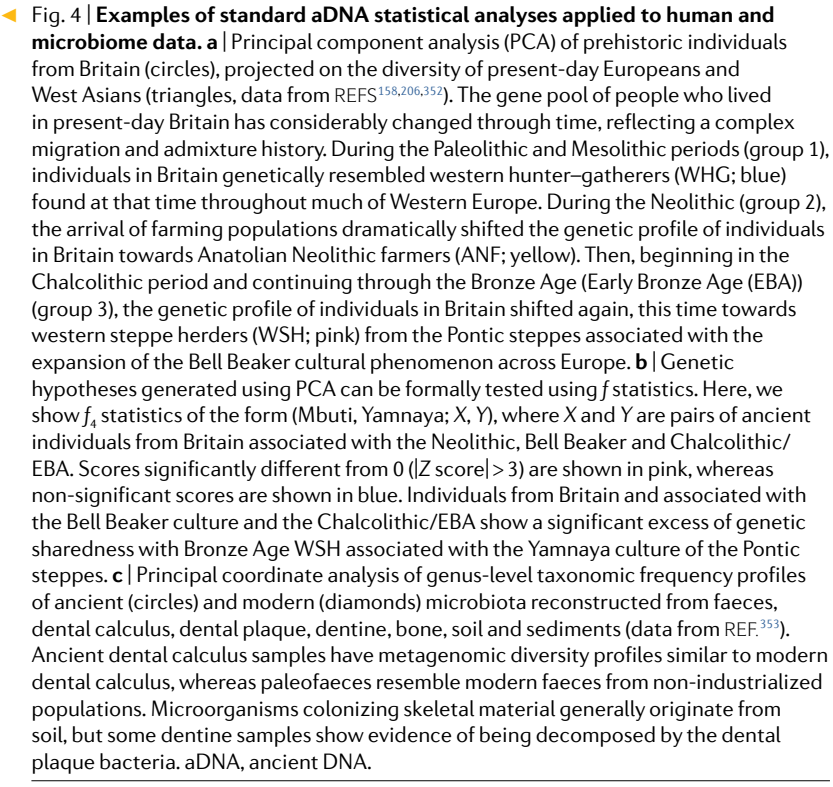


Fig. 4 | **Examples of standard aDNA statistical analyses applied to human and microbiome data. a** | Principal component analysis (PCA) of prehistoric individuals from Britain (circles), projected on the diversity of present-day Europeans and West Asians (triangles, data from REFS[158,206,352]). The gene pool of people who lived in present-day Britain has considerably changed through time, reflecting a complex migration and admixture history. During the Paleolithic and Mesolithic periods (group 1), individuals in Britain genetically resembled western hunter–gatherers (WHG; blue) found at that time throughout much of Western Europe. During the Neolithic (group 2), the arrival of farming populations dramatically shifted the genetic profile of individuals in Britain towards Anatolian Neolithic farmers (ANF; yellow). Then, beginning in the Chalcolithic period and continuing through the Bronze Age (Early Bronze Age (EBA)) (group 3), the genetic profile of individuals in Britain shifted again, this time towards western steppe herders (WSH; pink) from the Pontic steppes associated with the expansion of the Bell Beaker cultural phenomenon across Europe. **b** | Genetic hypotheses generated using PCA can be formally tested using *f* statistics. Here, we show f_4 statistics of the form (Mbuti, Yamnaya; *X*, *Y*), where *X* and *Y* are pairs of ancient individuals from Britain associated with the Neolithic, Bell Beaker and Chalcolithic/EBA. Scores significantly different from 0 (|*Z* score| > 3) are shown in pink, whereas non-significant scores are shown in blue. Individuals from Britain and associated with the Bell Beaker culture and the Chalcolithic/EBA show a significant excess of genetic sharedness with Bronze Age WSH associated with the Yamnaya culture of the Pontic steppes. **c** | Principal coordinate analysis of genus-level taxonomic frequency profiles of ancient (circles) and modern (diamonds) microbiota reconstructed from faeces, dental calculus, dental plaque, dentine, bone, soil and sediments (data from REF.[353]). Ancient dental calculus samples have metagenomic diversity profiles similar to modern dental calculus, whereas paleofaeces resemble modern faeces from non-industrialized populations. Microorganisms colonizing skeletal material generally originate from soil, but some dentine samples show evidence of being decomposed by the dental plaque bacteria. aDNA, ancient DNA.

that implements an explicit genetic model aimed at partitioning individuals into ancestries, fitted to genotype frequencies[214–216], that has now been extended to model the temporal structure of the data[217]. However, these approaches are sensitive to (and thus can be biased by) genetic drift between time periods and differences in relative sample size between ancestries. This contrasts with methods based on *f* statistics[217–219], which leverage covariation of allele frequencies between populations and can retrieve unbiased admixture proportions. Although more cumbersome than clustering approaches for initial data exploration, *f* statistic-based methods can, in many cases, also provide a statistical test for proposed ancestry models.

Three individual statistics are central to the *f*-statistic framework. f_2 statistics measure distances between pairs of populations and/or individuals. f_3 statistics can reveal unambiguous evidence for situations in which one individual (or population) is formed by the admixture of two others[219] but can also provide a measure of shared genetic drift when an outgroup is included[220]. Finally, f_4 statistics test for asymmetries between populations that are indicative of gene flow[219] (FIG. 4b). If a five-population history of a specific topology can be assumed, a ratio of two f_4 statistics can be used to estimate ancestry proportions unbiased by genetic drift[219]. In more generalized developments of the *f*-statistics approach in the tool qpAdm (and qpWave), the predicted *f* statistics of a one-source[221] or two to four-source[158,210] ancestry model can be obtained if a set of reference populations or genomes can be posited that are more distantly related to the ancestry sources than populations defined by users to represent sources. This approach then provides a *p* value for the ancestry model tested and also, for valid models, an unbiased ancestry proportion estimation.

Genotype imputation. With some exceptions, ancient genomes and genome-wide enrichment of SNPs are sequenced at limited depth of coverage, which precludes the determination of genotypes. Statistical inference of missing genotypes is possible using a process called genotype imputation, which assumes that the distribution of haplotypes present in the population is known and that sufficient coverage is available, typically around 1× (REF.[222]). In practice, the haplotype distribution is often approximated from large-scale reference panels such as the Haplotype Reference Consortium[223] for human populations. This may be adequate for relatively recent historical time periods[224] and/or areas such as Europe where many genomes from modern human populations have been collected[222] but may be underpowered in cases of limited reference panels, as is true for many populations from elsewhere in the world[225]. Imputation can help genotype single loci — such as lactase persistence[226] — to full genomes[52,222,227] and estimate runs of homozygosity and inbreeding coefficients from low-coverage data[227]. Promisingly, new approaches relying on linkage information from reference panels of modern haplotypes that are compatible with a minimum of 0.3× coverage data are now emerging[228]. Patterns of non-random allele association within haplotypes can also improve the inference resolution of population structure and outperform methods using unlinked variation[222,229]. Nevertheless, imputation can also infer false genotype calls[230], and so care must be exercised when using imputation to investigate the evolutionary history of specific traits.

Microbiota profiling. In addition to containing genomic fragments from the focal species, aDNA extracts also generally play host to an entire metagenomic diversity of environmental microorganisms that mainly colonized the subfossil material after death[7,231]. The presence of often dominant non-host DNA templates within aDNA libraries can considerably increase genome sequencing costs but can be filtered out using target-enrichment techniques or enzymatic digestion of over-represented bacterial sequence motifs[7]. In studies of pathogens (reviewed elsewhere[13]) and oral, gastric and faecal microbiota, however, the metagenomic content itself can be the intended research target if pathological lesions, dentine, dental calculus[26], stomach contents[231,232] or coprolites[233–235] are analysed.

Taxonomic profiling of library metagenomic content is typically carried out using a sequence identity threshold against comparative databases (such as MALT[236]; FIG. 4c) or *k*-mer representativity patterns (for example, Kraken[237]), which break down the frequency distribution of each sequence motif of a predefined length of *k* nucleotides. The sensitivity and specificity of taxonomic assignment largely depend on the representation characteristics of the comparative database but are robust to post-mortem DNA damage[238,239]. Curated databases of microbial markers such as MetaPhlAn2 (REF.[240]) ensure specificity but may lack sensitivity for environmental

16S meta-barcodes
Selected variable regions of the 16S ribosomal RNA gene whose sequence provides taxonomic resolution amongst bacteria and archaea.

DNA methylation
A biological process by which the activity of a DNA segment is modified without changing the underlying sequence but by adding methyl groups to the DNA molecule.

Bisulfite conversion
A chemical reaction using sodium bisulfite that converts unmethylated CpG dinucleotides into UpGs but leaves methylated CpGs intact, thereby allowing the detection of DNA methylation by sequencing.

Immunoprecipitation
A molecular laboratory technique by which specific molecules are purified on the basis of their chemical affinities for particular protein groups, such as antibodies.

and/or non-human associated microbial taxa, whereas large uncurated sequence repositories such as the National Center for Biotechnology Information (NCBI) nucleotide database may lack specificity (reviewed elsewhere[27]). Several bioinformatic pipelines automating analyses from mapping to statistical profiling are available, such as HOPS[241] and metaBit[242]. In addition, SourceTracker[243] and CoproID[233] can estimate mixture proportions from known candidate sources, such as soil, gut and oral microbiota. Authentication of microbial taxonomic profiles is difficult and requires cross-validation through different software[238] and analyses of specific sequence characteristics, including through mapping against the genomes of candidate species and recovery of post-mortem signatures of molecular degradation[27].

Although 16S meta-barcodes are extensively used for profiling modern microbiota, the approach is impracticable on ancient material because of amplification biases introduced during PCR owing to extensive DNA fragmentation[133]. 16S ribosomal RNA gene amplification is designed to target hypervariable regions in order to distinguish microbial taxa, and these sequences are clustered based on similarity (≥97%) with bacterial taxa. However, most aDNA sequences are smaller than 200 bp in length, the minimum for 16S ribosomal RNA variable region amplification. For example, a 2015 study[133] showed that amplified libraries of the 16S V3 region did not accurately reflect microbial taxa within aDNA samples when compared with shotgun metagenomic data. This was, in part, owing to biased amplification of length polymorphisms whereby taxa with deletions were over-represented and taxa with insertions were under-represented[133]. Shotgun metagenomic approaches are thus recommended.

DNA methylation. Two main experimental methods have been applied to map DNA methylation signatures on ancient genomes. The first builds on methods commonly used on fresh DNA, such as bisulfite conversion[244,245] and immunoprecipitation[62,167]. Bisulfite conversion uses sodium bisulfite to convert unmethylated cytosines into uracils, similar to the C to U conversion that occurs in post-mortem DNA damage. However, bisulfite conversion is also extremely harmful to DNA and, thus, generally not recommended for aDNA, despite a few successes on limited numbers of samples[244,245]. Immunoprecipitation is likewise extremely limited in practice owing to the fast post-mortem decay of CpG dinucleotides[167] and DNA fragmentation[62], which reduce the number of potential DNA methylation targets available per fragment.

The second category of methods for aDNA methylation analysis relies on indirect statistical inference based on patterns of post-mortem DNA decay that are ubiquitous in plants[246] and animals[28,29]. The idea is implemented in open-source statistical packages such as *epiPALEOMIX*[247] and *DamMet*[248]. It leverages NGS data generated following USER treatment, which maintains C to T misincorporations only at methylated sites[29]. Although obtaining reliable estimates at single-nucleotide resolution requires impractical sequence coverage (≥80× coverage[248]), the methodology has been applied successfully to assess DNA methylation levels within genomic regions.

Applications

The application potential of aDNA analysis has a broad scientific scope and relevance to key archaeological, ecological and evolutionary questions. The following section highlights some of the areas that have received heightened scholarly attention, including human population history, plant and animal domestication, and the origins and evolution of pathogens and microbiomes, as well as the impact of global climatic changes on past faunal and floral communities.

Human genomics. No area of aDNA analysis, except that of microbial archaeology[13], has benefited as much from NGS as that of past human population genetics (FIG. 5). The wealth of newly available ancient human genomic data has been instrumental in transforming our understanding of the human past, from a previous narrative of long-term population continuity and isolation to one in which mobility and population mixture have played much more prominent roles[11].

The first ancient human genome was sequenced in 2010 from the hair of an approximately 4,000-year-old Saqqaq Paleo-Inuit man from Greenland[6]. This made the possibility of sequencing of authentic ancient human genomes a reality, something that was until then considered unlikely owing to pervasive sequence contamination in the pre-NGS era[249]. The draft genomes of two archaic hominins, the Neanderthals[7] (1.3× coverage, composite of 3 individuals) and Denisovans[15] (1.9× coverage), were released shortly after the high-coverage Saqqaq Paleo-Inuit genome[6] (20× coverage). Both archaic genomes revealed evidence of admixture amongst various hominin lineages[8], in which Neanderthals and Denisovans contributed significant ancestry to modern non-African populations and modern populations of Australasia and Oceania, respectively[221], contributing to potential phenotypic and health consequences today[250].

Since 2010, ancient human genomics has moved forwards with incredible speed, and current publications typically include genome-scale data and/or whole-genome sequences from dozens to hundreds of individuals[5,251–253]. Current understanding of genomic variation within pre-Holocene populations is lagging behind that of more recent populations owing to the scarcity of the fossil record and relatively limited DNA preservation, although partial nuclear genomes from pre-Neanderthal groups[193] and genomes from multiple Upper Palaeolithic modern humans[37,254–257] have been successfully recovered. Europe, and more generally west Eurasia, has been the focus of most research[258], although significant progress has also been made in understanding the genomic history of other regions, such as East Asia[257,259], the Americas[6,220], Oceania[190,260], Southeast Asia[261,262], the circum-Mediterranean[263] and Africa[264]. These other regions include climatic zones that were previously believed not to be suited for long-term DNA preservation.

Studies of west Eurasia have revealed that the genetic diversity of present-day Europeans is primarily formed

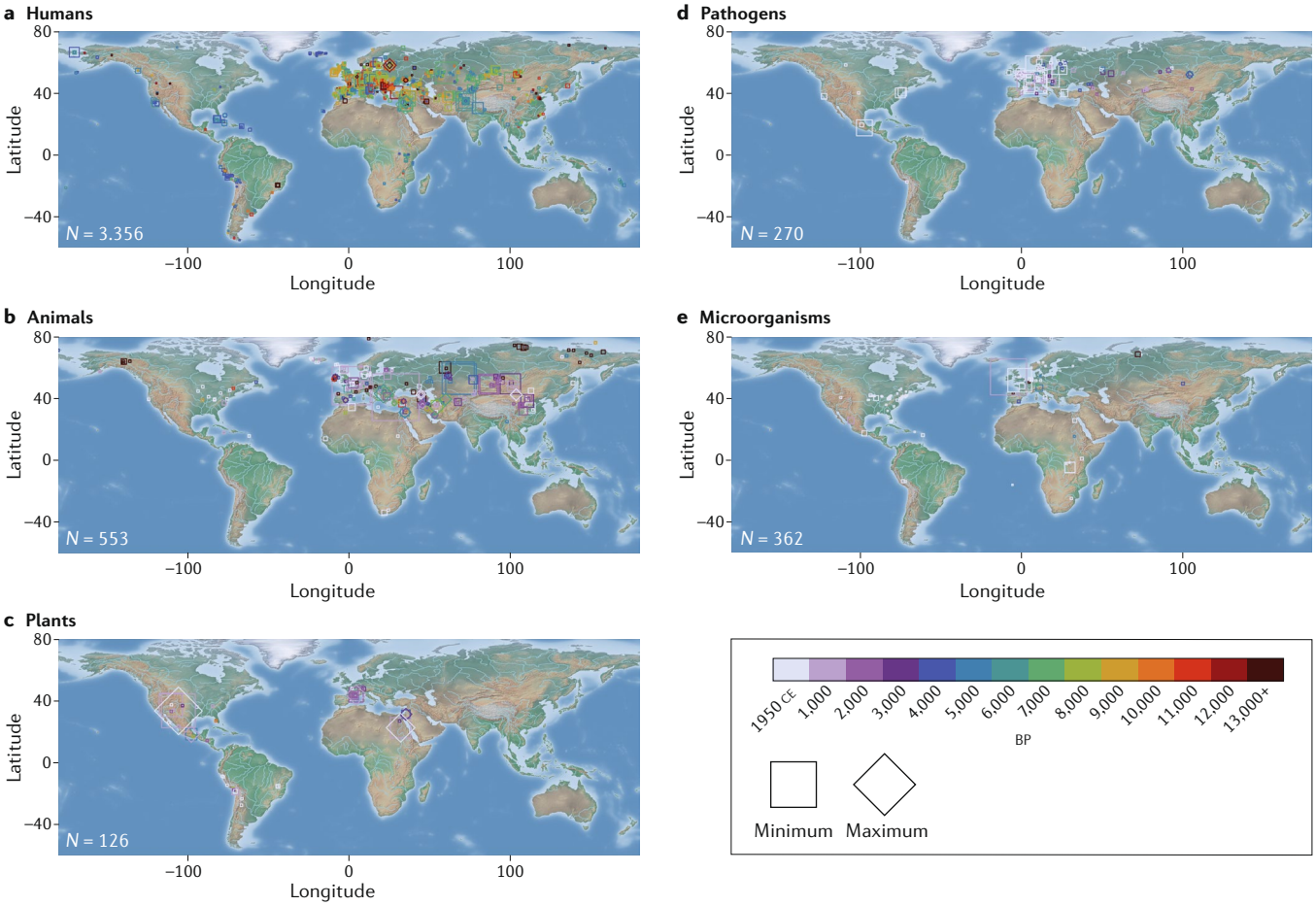


Fig. 5 | **Geographical and temporal distribution of ancient specimens analysed at the genome scale. a** | Humans (data from Ancient Human DNA uMap). **b** | Non-human animals. **c** | Plants. **d** | Pathogens (data from REF.[353]). **e** | Microbial metagenomes (data from REF.[353]). The minimum (square) and maximum (diamond) age of the samples analysed at a given location are provided. Only squares are shown in case the temporal range overlaps only one time interval. Colours correspond to approximately 1,000-year time intervals, except for the oldest that includes genomes dating to 13,000 years ago or longer, including Late and Middle Pleistocene specimens. Size of squares and diamond is proportional to the number of samples analysed at a given location but is adequately scaled in each panel for clarity and according to the total number of specimens (*N*) for which both geographical and temporal data are available. BP, before present (that is, prior to 1950 CE).

by three major ancestry components[210], consisting first of early hunter–gatherers[194,265] who later admixed with farmers leaving Anatolia approximately 9,000 years ago, and who brought with them agricultural innovations such as domesticated crops and livestock and introduced the Neolithic lifestyle into Europe[266]. A third genomic component was later overlaid following the migration of pastoralists culturally associated with the Yamnaya horizon from the Pontic steppe around 5,000 years ago[158,226] (FIG. 4a,b). It is now believed that it was this pastoralist migration, and not earlier Neolithic population movements, that likely brought the proto-Indo-European language to Europe[158,226]. This was then followed by the westward expansion of the Bell Beaker complex that is genetically associated with an almost total population replacement of groups in Britain[206].

Further to the east, archaeogenomic studies have shown that Central and Inner Asia underwent a population history very different from that of Europe, even though they were impacted by some of the same events. These areas were initially populated by groups of hunter–gatherers who were among the first to domesticate the horse[251] and who were related to contemporary Native Americans. However, in contrast to the hunter–gatherers of Europe, these peoples did not experience migration from Anatolia and never underwent the same processes of Neolithization[253]. Nevertheless, much of Asia has witnessed major migrations since the Bronze Age[251–253,267], including from Yamnaya-related and subsequent western steppe populations who spread pastoralism across the continent. These population dynamics resulted in Central and Inner Asia shifting from initially being occupied by groups of peoples related to Native Americans, to being occupied by peoples of mixed western Eurasian genetic ancestry and then, finally, to becoming inhabited by peoples with greater Northeast and East Asian ancestry. In the process, the language topology of the region also changed from being largely Indo-Iranian to becoming predominantly Turkic or Mongolic today[252,253].

Beyond Eurasia, studies of ancient human genomes have also substantially contributed to our understanding

Population replacement
A population process by which the gene pool of one local population is at least partially replaced by that coming from another, genetically distinct, population.

of the population history of the Americas, from its initial peopling and the subsequent dispersal and diversification of indigenous populations[268–270] to the impact of European contact and colonialism[271]. Recent studies are also informing the population history of tropical regions previously thought to be beyond the reach of aDNA owing to extensive DNA degradation conditions, such as the Caribbean, which now appears to have been colonized in at least three waves from migrants originating in both North and South America[272]. The study of ancient human genomes has also contributed to the reconstruction of events for which the historical record is limited or intentionally concealed. One such example is the application of ancient genomics to the study of the Transatlantic Slave Trade, which has allowed the identification of specific places of origin of African-born enslaved individuals who were forcibly taken from Africa to the Americas, revealing the diversity of their ethno-geographical origins within West Africa and beyond, as well as aspects of their lives in the Americas[273,274].

Domestication. The domestication of plants and animals represents a key stage in human history and has long been an area of unresolved debate about the timing, location and number of domestication sources[275]. Pre-NGS era studies identified key loci under early selection in crops such as maize[69] and pigs[276], providing glimpses into the process as well as into their likely routes of spread. Now, domestication models are being rewritten as genomic data reveal a dynamic and complex history that includes complete population turnover, adaptive introgression between wild and domesticated forms, an unexpected temporal and spatial distribution of origin and selection, and a larger role for natural dispersal processes (see[12,277] for recent reviews on these topics).

Large-scale population replacement occurred in dogs[278], wolves[196,279,280], horses[148,281], pigs[282] and crops such as potato[283]. For example, indigenous American dogs were completely replaced by later European dogs[278] with some introgression from Inuit dogs[284], and modern domestic horses did not arise from the Eneolithic Botai culture of central Kazakhstan as previously thought[281] but from a later Bronze Age expansion[148]. By contrast, a staggering level of constancy is evident in grapes with more than 900 years of uninterrupted vegetative propagation[72].

Adaptive introgression has emerged as a recurrent feature in response to past environmental dynamics. For example, maize assimilated wild adaptive variation enabling northward expansion[285], as did flax in Europe[286]. Cattle were introgressed by wild aurochs[287] and with zebus, thereby gaining alleles adaptive for response to drought[288]. Such findings overturn the concept of domestication as a process of isolation between wild and domesticated populations, and suggest instead complex instances of gene flow, as is evident in the mosaic ancestry of ancient goats[289] and in the multiple emergence events of semi-domesticated maize varieties[70,74], which were later fully domesticated at secondary centres in a stratified domestication process[290]. Paleogenomic data also do not support the presence of a strong demographic bottleneck during early domestication stages (see[291] for a review in major crops) but, rather, studies of wheat[292], sunflowers[293] and horses[148] support more recent losses of genetic diversity.

Finally, time-stamped genomes have enabled the temporal stratification of both phenotypes and selective regimes. In sorghum, early selection was related to plant architecture and only later switched to increased sugar metabolism[294]. In maize, complex phenotypes such as days to flowering have been reconstructed from genomes showing latitudinal adaptation to day length[295]. Early selection of coat colour is evident in pigs[282] and goats[289], but some traits were acquired later, such as behaviour-related and productivity traits in chickens during the medieval period[296,297] and morphotype and speed in horses[148,298]. Importantly, damaging mutation load was a feature of later selective episodes and breeding but not part of the initial domestication of horses[298], maize[295] and sorghum[294]. Continuing to characterize past genomic variation in ancient crops and animal breeds will be instrumental for developing the future sustainability of agriculture.

Pathogens and microbiomes. Together with classical approaches in paleopathology and paleodemography, aDNA from microorganisms, including pathogens and commensals, can provide insights into the health of ancient peoples as well as shifts in diets and disease ecology. Initial studies of ancient microorganisms used PCR-based methods to identify specific pathogens (such as those causing skeletal lesions characteristic of tuberculosis and leprosy)[299], to analyse the first historic pathogen genome from the 1918 influenza virus[300] and to explore the ancient microbiome[27]. However, such approaches suffered from an inability to distinguish ancient and modern contaminating microbial DNA[27,301]. More recently, NGS-based methods have resulted in the successful recovery of authenticated pathogen genomes from not only *Mycobacterium tuberculosis* and *Mycobacterium leprae* but also many pathogens that leave no visible evidence in the skeleton, such as *Yersinia pestis*, *Helicobacter pylori*, *Vibrio cholera*, *Salmonella enterica* serovar Paratyphi C, variola virus, human parvovirus and hepatitis B virus (reviewed elsewhere[13]). For example, analyses of the victims of both the Plague of Justinian and the Black Death have shown that *Y. pestis* was indeed the cause of both pandemics[65,302], and further genomic analysis has provided insight into the course of these pandemics over time, including parallel changes in the pathogen itself[303–305]. Surprisingly, extinct strains of *Y. pestis* have also been identified in ancient individuals dating from as early as the Late Neolithic and Bronze Ages throughout Eurasia, and genomic data show the progression of evolutionary changes that have affected its virulence through time, including the acquisition of a transmission mode via fleas (reviewed elsewhere[13]).

Although the analyses of ancient pathogen genomes have shown plague to be older than expected, they suggest that the origins of the tuberculosis complex (*M. tuberculosis* and related strains) in humans may have evolved more recently (~3,000–6,000 years ago)[306,307]. PCR data exist that putatively identify tuberculosis in older individuals, but genomic data from bone samples

that would distinguish contaminating microbial DNA are not yet available[27,301]. Additional genomic data from ancient tuberculosis strains could show that the current estimate of time to the most recent common ancestor reflects a bottleneck or other population dynamics in this clonal pathogen. The ancient tuberculosis data also show the significance of pathogen exchange among animals and humans. In South America, before European colonization, aDNA has shown that people were affected by tuberculosis strains that are most closely related to *Mycobacterium pinnipedi*, a zoonotic form of the pathogen that usually infects seals and sea lions. This strain was likely transmitted from South American seals to humans through consumption of undercooked seal meat or during butchering, after which human to human transmission may have begun[306].

Similarly, NGS has facilitated microbiome analyses from a range of ancient contexts including dental calculus, coprolites, latrine sediments and mummies[308,309]. Comparisons between modern and ancient oral microbiota have suggested possible shifts in relation to increasingly carbohydrate-rich diets during the Neolithic and Industrial revolution[44], although further work revealed the need to account for other possible factors driving oral microbiota profiles, such as the biofilm maturation stage[310], tooth type and surface[311]. Likewise, studies of the ancient gut microbiome have revealed previously unknown microbial diversity in the human gut, as well as the loss of key microbial symbionts in present-day industrialized populations[85,235]. The microbiome is also the source of several recently evolved human pathogens, including the causative agents of diphtheria (*Corynebacterium diphtheriae*), gonorrhoea (*Neisseria gonorrhoeae*), bacterial meningitis (*Neisseria meningitidis*) and pneumonia (*Streptococcus pneumoniae* and *Haemophilus influenzae*) (reviewed elsewhere[309]), and thus understanding its evolution and changing ecology through time is critical to understanding infectious human disease.

Extinction and climate change. aDNA has fed debates not only on the phylogenetic placement of many extinct species but also on the causes of their disappearance. This can be investigated by correlating known human activities or climatic events with population expansions and declines as revealed from serially sampled DNA data. Dynamic population size trajectories can be estimated from calibrated gene genealogies within a serial coalescent statistical framework (reviewed elsewhere[312]). Additionally, the regions showing paleoclimatic conditions compatible with the presence of a given species can be inferred from spatio-temporal fossil distribution data and bioclimatic data. Key to all such analyses is the availability of radiocarbon dates for the specimens analysed using genetic techniques. Radiocarbon dating should thus be highly recommended wherever possible.

Evidence currently available in bison, horses, reindeer, musk oxen, woolly mammoths and woolly rhinos suggests that climate change is a common driver of population size changes but also that species show individualistic responses in the face of climatic and anthropogenic pressure[20]. Human activities have been found to possibly mediate species demise through various mechanisms, including direct over-exploitation, as shown in moas[313] and great auks[314]. Extinctions can also have many indirect causes that can be difficult to disentangle, such as human-mediated habitat disturbance or introduction of pathogens or predators. For example, the occupation of caves by Aurignacian humans competed with the natural homing behaviour of cave bears for places in which they could hibernate to survive the winter[23,24]. As regards the extinct huia birds of New Zealand, the drastic reduction of forest cover and the translocation of mammal predators during the European settlement may have been fatal to these endemic passerines[315].

Complete genome data from single diploid individuals have provided new avenues for reconstructing paleodemographic trajectories using the pairwise[316] and multiple[317] sequentially Markovian coalescent modelling frameworks. This approach has revealed that archaic hominin populations survived extinction over long periods of time despite highly reduced effective population sizes[8–10]. Analysis of the complete genome from a late-surviving woolly mammoth has identified two episodes of severe demographic collapse in the Early-Middle Pleistocene and at the Pleistocene–Holocene transition[3]. The resulting reduced genetic diversity and the accumulation of numerous potentially deleterious mutations may have precipitated the population's extinction through the alteration of important functions in development, reproduction and olfaction[17–19]. Demographic reconstructions can, however, be biased by strong natural selection acting on genomes, especially in extreme cases of large population sizes. This was observed for the passenger pigeon that went from a census population size of several billion to extinction in only decades[318]. By reconstructing temporal baselines prior to environmental and/or human impact, it is increasingly clear that aDNA from museum and archaeological remains can advantageously complement current genetic diversity assessments to establish extinction risks and conservation priorities[319], as previously hinted for Przewalski's horses[320] and arctic foxes[321].

Environmental DNA. The retrieval of animal and plant DNA from sediments, ice and lake cores, commonly known as environmental DNA (eDNA), is probably the only research area that traces its origins to the field of aDNA itself rather than being adopted from contemporary genetics. Early pioneering work successfully retrieved DNA fragments from extinct mammals and birds from only a few grams of Siberian permafrost and New Zealand cave sediments[90]. If requiring large drilling equipment, sampling procedures should be experimentally tested for contamination, for example by introducing synthetic plasmids of known sequences at the surface of the equipment whose penetration into the drilling core can be tracked[22,48]. The experimental workflow typically includes DNA extraction and the sequencing of biomarkers providing taxonomic resolution[322], and special attention must be given to rule out on-site contamination[48] and stratigraphic leaching[323]. Early analyses relied on molecular cloning of PCR-amplified

Environmental DNA
(eDNA). Fragments of DNA that are preserved within sediments and water that can be used for a fast, cost-effective monitoring of the ecology of a given region.

Stratigraphic leaching
The migration of DNA across strata in sediments caused by water movement, microorganism growth or bioturbation and compromising the reliability of the stratigraphy, that is, the order, position and age of the geological layers formed by the different piles of sediments.

meta-barcodes, but massively parallel sequencing facilitated access to the entire molecular diversity of individual animals and plants[324]. Current approaches build on shotgun metagenomics[48] and more economical techniques involving target enrichment[49,325].

One of the great benefits of eDNA is that it allows the detection of all domains of life, from microorganisms to vertebrate species alike, even in the absence of macrofossils[325,326]. Taxonomic resolution is, however, limited owing to the extensive fragmentation of eDNA, and varies with the completeness of comparative sequence databases. The extent to which quantitative assessments can be obtained is currently unknown, although new biomass proxies aim at this objective[136]. The exact sources of eDNA fragments and the conditions governing their preservation remain largely unknown. eDNA is assumed to derive from skin cells, faeces, urine and microfossils, and various mineral particles have been suggested to favour its preservation within sediments[90]. Microscopic bone and tooth fragments that are too small to be identified, and are thus confused with sediments, have also been speculated to potentially represent a significant source of eDNA[49].

eDNA has proved a powerful approach to assess the first and last appearance of taxa in the fossil record, and has been applied to investigate the timing of the extinction of woolly mammoths in mainland Alaska[326] and Yukon[325] and the survival of spruce in Scandinavian refugia during the last Ice Age[130]. eDNA has also provided insights into the spatial and temporal distribution of animal species, including extinct hominins[49], and the impact of global climatic change[22,48,90,91,327]. All of these studies have in one way or another challenged the validity of climate niche predictions and commonly accepted models of extinction, biogeography and archaeology.

Reproducibility and data deposition

Possible confounding factors. Given that numerous tools and approaches are available for analysing ancient genomic data, reproducibility of results can be hampered by the lack of sufficiently detailed descriptions of each analytical step in the methods sections of publications. Although some guidelines exist for the analysis of ancient genomic data — for example, for mapping[162,181,183] or data authentication[56,57,328] — and automated tools embedding critical analytical procedures within open-source pipelines are available[187,188,241,242], there is still no consensus on a 'gold standard' method for carrying out (or reporting) most analyses. More often, the selection of programs and parameters is research group-based, with publications reporting different analytical approaches and criteria. The list of factors to consider extends far beyond the minimal data requirement for specific analyses[66] and includes platform-specific sequence error profiles[169]; the magnitude of post-mortem DNA damage and contamination, and their impact on downstream inference[55]; base quality rescaling[57]; read alignment parameters[180,182]; database selection and taxonomic classifier parameters[27,238]; awareness of the applicability conditions of statistical methods; and many more. In the face of the complexity and diversity of possible confounding factors, it is of the utmost importance that all relevant analytical parameters, including software versions and parameters (even if default), are described in full detail in the methods section of any scientific article reporting aDNA data.

Public repositories. To be accepted as valid forms of evidence, raw sequence data and alignments against reference genomes must be made freely available through public repositories such as the Sequence Read Archive (SRA) and the European Nucleotide Archive (ENA) in order to ensure research reproducibility and future analyses. Entirely curated data sets are also made available upon publication through individual laboratory websites (for example, the Reich laboratory website). Full traceability of the underlying experimental procedures, including sequencing chemistry and base calling software, is necessary to account for possible data structure deriving from technical artefacts. We recommend that raw, unfiltered sequencing data be uploaded as compressed fastq files together with the alignment files underlying the analyses presented — in BAM format for individual (mito)genomes and in multifasta format for barcode alignment or phylogenetic reconstruction. Labels — for example, read groups and sample names — should match those provided in the supplemental sections of the original publication.

Long-term legacy. The destructive sampling underpinning ancient genomic work has raised severe concerns pertaining to the long-term viability of the data produced. Current approaches are mainly based on whole-genome shotgun sequencing or targeted SNP capture. Although the latter is both time and cost-effective[160], and remains the only methodology showing sufficient sensitivity to gather data from extremely degraded specimens[119], it considerably restricts the amount of information retrieved from samples showing better molecular preservation. In ancient human genome studies, this approach typically queries 1.2 million positions known to be polymorphic amongst a worldwide panel of contemporary human populations[160], leaving other positions and other DNA material present in the extracts, such as those from pathogens or microbiota, unexplored. Whole-genome shotgun sequencing is instead not limited to any pre-selected target regions or genomic variants but, rather, aims to uniformly cover the entire metagenome. Although more expensive to generate and more demanding to analyse, this untargeted approach can reveal new informative variants and produce data that are applicable to future research questions. Regardless of the approach, however, it is also possible to create from the original extracted aDNA an immortal DNA sequencing library that can serve as a long-term genetic archive of the sample, and which could be stored in a museum or laboratory cryo-facility. If prepared correctly, this library can be theoretically reamplified indefinitely without exhaustion or loss of complexity[143]. We advise curators to balance the uniqueness of the material considered, known or estimated DNA preservation rates, experimental costs and long-term plans for both DNA library and data archiving when authorizing destructive sampling.

Limitations and optimizations

Wet-laboratory methods. Owing to the sensitive nature of aDNA, preparation and manipulation of aDNA must be carried out in laboratory facilities with positive air pressure, UV surface irradiation and strict cleaning procedures to ensure minimal contamination. Clean laboratory facilities must be physically separated from other laboratory areas where DNA is amplified, captured and/or sequenced. Samples must be selected according to the main objectives of the study. For example, whereas petrosal bones are generally preferred for retrieving genome-scale data from host species, their metagenomic potential for pathogen studies is limited[66]. Rates of aDNA preservation also vary at microscopic scales. Therefore, multiple analyses repeated on small amounts of material can increase the chances of success compared with a single analysis from a large sample, which averages out preservation rates and maximizes destruction. No DNA analyses should be carried out on precious, rare material in the absence of preliminary analyses, suggesting feasibility and supporting molecular preservation on site. Specific chemical treatment of bone powder, for example with bleach[40,42], can reduce the exogenous DNA fraction and may be attempted in cases of repeated failure due to contamination; however, such treatment can also artificially age modern contaminant DNA, and so authentication of bleach-treated samples should be performed with care. DNA libraries should contain multiple indices to eliminate the inclusion of spurious recombining DNA templates in downstream analyses[142]. PCR amplification of DNA libraries must be carried out in conditions maximizing molecular complexity, and no DNA extracts and/or DNA libraries should be exhausted for a single analysis. In the case of target enrichment, two successive rounds of capture generally augment on-target recovery rates[166]. Library pooling prior to sequencing helps to reduce sequencing costs but requires that the base composition is balanced, especially for Illumina platforms, as a balanced base composition within the first six nucleotides is instrumental for calibrating the fluorescence measurement of nucleotide bases. This can be achieved by spiking library pools with sufficient amounts of a PhiX DNA library prior to sequencing. Large-scale projects can benefit from nascent automation procedures both for the preparation of DNA libraries and for their capture[49,126,144]. In all circumstances, the experimental procedures implemented from sampling to sequencing must be recorded and fully described (and, ideally, shared through online repositories such as protocols.io) so as to assess the possible impact of experimental differences in the sequence data.

Dry-laboratory methods. With genome-scale sequence data, the number of possible statistical analyses is virtually limitless. Common caveats include the impact of contamination and post-mortem DNA damage and of read alignment parameters. For highly degraded or contaminated material, analyses may be conditioned on reads showing evidence of post-mortem DNA damage[55,189,192] and/or limited to transversions[7] so as to mitigate both rampant contamination and damage-related nucleotide misincorporation. This approach, however, significantly reduces the amount of data available. Quantitative assessment of damage and contamination followed by explicit simulation[180] can help establish the robustness of the conclusions. Microbial taxonomic assignment should be cross-validated through multiple approaches and the presence of potential pathogens must be established by checking that the read edit distance distribution against the reference genomes of close relatives indicates closer proximity with the candidate[27]. The amount of sequence data necessary for a given analysis depends on the underlying population history and genetic diversity present in a species. For example, a minimal threshold of 10,000–30,000 SNPs is commonly used for assessing genetic relationships of human individuals at intra-continental scales[329,330] but molecular sexing of mammal species and identification of first-generation hybrids generally requires no more than 1,000–10,000 mapped reads[200]. The computational procedures must be fully described with explicit reference to analytical parameters so as to ensure reproducibility (and, ideally, shared through online repositories such as GitHub or Bitbucket).

Outlook

Working together. Within the past decade, aDNA research has come of age and has complemented the toolkits of archaeologists and evolutionary biologists, with techniques providing an unprecedented resolution of detail on past environments, societies and individuals. Perhaps the most important contribution of aDNA is that it offers access to a wealth of biological information that can help shape and test working hypotheses about the past in a quantitative manner. Adopting such a hypothesis-testing framework comes naturally for evolutionary biologists who share similar methodologies, including phylogenetic reconstruction and population genetics. This approach, however, differs from that of archaeologists, whose research is primarily grounded within the epistemological frameworks of the social sciences and humanities. As a fast-evolving and inherently interdisciplinary field, aDNA research is where these distinct disciplinary approaches mix and meet — to both great effect and occasional mutual misunderstanding. For example, most aDNA work to date has focused on big picture studies comprising large temporal and spatial scales. Although necessary for delineating major trends and new working hypotheses, such studies have sometimes divided the archaeological community, as they necessarily oversimplified the inherent complexity of past societies and the archaeological record itself. Moreover, the often unreflected use of problematic terms and the simplistic equivalence of genetic and cultural groupings (BOX 1) have found major criticism[104,107]. Whether seemingly rapid shifts in ancestry detected in aDNA data were the result of sudden massive population replacements and/or long-term individual human mobility owing to institutions such as patrilocality has also been a matter of discussion and debate[331,332]. However, now that studies have created a basic overview of the population genomic developments in many regions of the world, it is possible to zoom into more local contexts. Future projects can, for instance, study

to what extent biological relatedness was decisive for the constitution of social belonging expressed, for example, through the joint burial of individuals in collective graves over time. Paleodemographic profiles augmented with aDNA evidence on the virulence of past pathogens, as well as paleopathological data, markers of violence and activity and dietary isotopic signatures, can help to establish the underlying causes of health changes in past societies. The reconstruction of ancient trade routes and networks represents another area that could greatly benefit from joint research efforts combining aDNA, archaeological fieldwork and remote sensing, for example by integrating genetic markers of migration and metallurgic signatures of material provenance into geographical information system data analysis. Encouragingly, in the past few years, scholars with different backgrounds in the fields of (bio)archaeology have increasingly learned to work together and have developed many initiatives to establish and strengthen the crucial interdisciplinary dialogue and mutual understanding that forms the basis for a future reflective bioarchaeology[104]. It has become clear that the full community of stakeholders (including descendant communities, archaeologists, biological anthropologists and museum curators) needs to be embraced, from sampling design to data interpretation, to avoid misinterpreting or over-interpreting the evidence and to collaboratively build a strong foundation for future work (BOX 2).

Future directions. In studies of humans, aDNA research has thus far been most successful when reconstructing ancestry profiles of past individuals and their population affinities. Future work may increasingly leverage patterns of linkage disequilibrium across sites to gain finer-scale genetic resolution. This will prove especially useful for identifying subtle genetic affinities not captured by unlinked SNP information alone[229] and when DNA preservation limits the amount of genetic information recovered. However, gaining further insights into the past will require molecular investigations beyond humans, their diseases and their domesticates, to include the full range of species and biomaterials preserved in the archaeological and paleontological record, from textiles and ceramics to sediments. Integrating DNA data with other proxies, including stable and radiogenic isotopes[112], microremains and small-molecule metabolites[333] and paleoproteomic data[113], will greatly improve our knowledge about the lives of past peoples including their diets[26,334], environments, mobility[335], craft activities[336], drug and medicinal use[337] and their health status and stress[26,338]. The successful application of such multi-proxy approaches to the study of the 5,300-year-old stomach contents of the Iceman has, for example, provided unprecedented resolution into the nutritional habits and food-processing methods during the European Copper Age[339].

An increasing body of work has shown that genetic data in tandem with archaeological and paleopathological data can illuminate social rules governing past societies, including marital patrilocality and social inequality[205], consanguinity[222], inbreeding avoidance[255] and care for individuals with genetic disorders. Recent case studies (for example, for the Early Bronze Age in Southern Germany[205] and the Neolithic in Ireland[222]) have shed new light on past social institutions and societal transformations by deciphering the complexity of households and individual mobility as well as the dependency of wealth and status on biological relatedness. Here, archaeogenetic studies have the potential to dramatically advance our understanding of past social structures and the interplay of biological relatedness, health and mobility with social practices and worldviews. Armed with ancient epigenetic markers, future aDNA work will likely inform us about age at death of past individuals[29,248] and the biological and medical consequences of their social status.

Given that all of these approaches are destructive and archaeological remains are finite, however, it is critical that resources are used with the aim of maximizing the amount of data collected and that efforts to develop multi-proxy methods compatible with minute sampling are explored[112,113]. Efforts to better understand post-mortem DNA degradation and to repair aDNA lesions[109] should be pursued to help further minimize destructive sampling. Such efforts would facilitate DNA analyses into the Upper Paleolithic and Middle Pleistocene, and possibly beyond the current record for the oldest genome around 560,000–780,000 years ago[2]. Increasing focus on environmental aDNA fragments preserved in sediments rather than in macrofossils will also provide additional opportunities for minimizing destruction while, perhaps, enriching and providing a fuller account of the taxonomic diversity of plant and animal communities in past ecosystems[22,34,48,49]. In addition, applying these approaches to marine and deep-sea sediments may increasingly inform us about the resilience of the ocean system to past global environmental change[340], and reveal ecosystems and past events in landscapes that became submerged during the

Box 2 | **Who conducts aDNA research?**

The field of ancient DNA (aDNA) research has diverse and heterogeneous disciplinary origins. Initial studies were largely led by biochemists and primarily focused on evolutionary questions relating to extinct mammals, but the field quickly diversified to also encompass anthropological and forensics questions. Today, evolutionary biology and anthropology have become the two main disciplinary homes of aDNA research, and there has been tremendous growth in the number of aDNA laboratories over the past decade. Housed in museums, research institutes and university departments, the size of these laboratories ranges from small laboratories serving a single principal investigator to large laboratories with multiple principal investigators supporting dozens of technicians, students and postdoctoral researchers. Funding is also highly varied, with some laboratories receiving core government funding whereas others depend entirely on third-party research grants. Moreover, funding agencies, grant types and funding levels also differ substantially across countries and disciplines, and such differences contribute to large disparities across different aDNA laboratories in terms of their resources and research capabilities. International collaboration and community-building, the establishment of enduring teaching and training networks and the formation of academic societies and research consortia are needed to leverage resources across the field, remove barriers to access for smaller laboratories and support the growth of the field as a whole. Initiatives such as the biannual International Symposium for Biomolecular Archaeology (ISBA) conference and the Standards and Precautions and Advances in Ancient Metagenomics (SPAAM) workshop aim to address some of these challenges, but additional collaborative efforts, such as lobbying of funding agencies and collective bargaining to make large in-solution probe sets available through commercial vendors, are needed.

Holocene[134,136]. Such studies promise to have a transformative impact on our understanding of the peopling of the Americas and island southeast Asia, as well as early human expansions out of Africa.

Over the past decade, short-read high-throughput DNA sequencing technologies have fundamentally changed the field of aDNA, but, even now, new sequencing technologies are on the horizon[341]. At first glance, long-read sequencers such as PacBio and Oxford Nanopore seem to have little relevance for highly degraded aDNA; however, concatenating aDNA fragments separated by spacers within single library templates may offer new opportunities for aDNA sequencing, including the direct detection of base modifications. In the longer term, this may even represent the only economical solution for future paleogenomic studies as short-read sequencing technologies are slowly phased out. New advances in de novo genomic and metagenomic assembly likewise hold promise for moving beyond reference-based mapping and towards a more accurate and complete reconstruction of ancient microbial genomes and metagenomes[238], including improved strain separation, genomic architecture reconstruction and the identification and recovery of novel genes. The past decade has shown that aDNA researchers can be extremely creative in developing innovative solutions to harness the full power of available high-throughput sequencing technologies. In the future, it will be essential to continue to adapt our research toolkits to incorporate new and emerging technologies as we strive to fully access the complete amount of information preserved in the fossil and sedimentary record.

New developments in inferring function from ancient genetic sequences are also afoot. Genome editing technologies such as CRISPR–Cas9 (REF.[342]), induced pluripotent stem cells[343] and miniaturized organoid systems mimicking simplified organs[344] now offer the possibility to investigate the functional consequences of virtually any change in the sequence of a given genome. The comparison of our genome and that of archaic hominins has, for example, provided a preliminary list of the genetic changes that made us humans, and attempts to understand their consequences on human brain development have already started[345]. Future research will likely increasingly rely on these technologies and genome time-series data of model organisms to assess the adaptive consequences of past genetic changes.

Over almost 40 years, aDNA research has taught us surprising things about the human story and life on this planet. It has moved forwards in parallel with technological and computational advances in the life sciences and has grown and matured in response to new developments in the social sciences and humanities. The discovery that a vast and invisible molecular past survives in the archaeological record has had a transformative effect on the field of archaeology and promises many new and unexpected findings to come. In the future, there is no doubt that our ability to increasingly detect and reconstruct the molecular archaeological record will enrich the human story and contribute to the transdisciplinary research endeavour of understanding our shared human past.

Published online: 11 February 2021

1. Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A. & Wilson, A. C. DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**, 282–284 (1984).
2. Orlando, L. et al. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).
This article presents the oldest sequenced genome to date, from one horse metapodial preserved in permafrost for 560,000–780,000 years.
3. Palkopoulou, E. et al. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* **25**, 1395–1400 (2015).
4. Barlow, A. et al. Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.* **2**, 1563–1570 (2018).
5. Margaryan, A. et al. Population genomics of the Viking world. *Nature* **585**, 390–396 (2020).
6. Rasmussen, M. et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
This article reports the first ancient human genome sequenced from the hair shaft of a 4,000-year-old Paleo-Inuit. The authors find evidence for genetic discontinuity with modern Greenlanders, supporting multiple migration waves into Arctic Greenland.
7. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
This article reports the first Neanderthal genome obtained from the DNA extracts of three paleontological bones, establishing the Neanderthal genetic legacy within modern human genomes and describing important wet-laboratory and dry-laboratory methodologies that have shaped the following decade of aDNA research.
8. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
9. Prüfer, K. et al. A high-coverage Neandertal genome from Vindija cave in Croatia. *Science* **358**, 655–658 (2017).
10. Mafessoni, F. et al. A high-coverage Neandertal genome from Chagyrskaya cave. *Proc. Natl Acad. Sci. USA* **117**, 15132–15136 (2020).
11. Nielsen, R. et al. Tracing the peopling of the world through genomics. *Nature* **541**, 302–310 (2017).
12. Frantz, L. A. F., Bradley, D. G., Larson, G. & Orlando, L. Animal domestication in the era of ancient genomics. *Nat. Rev. Genet.* **21**, 449–460 (2020).
13. Spyrou, M. A., Bos, K. I., Herbig, A. & Krause, J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat. Rev. Genet.* **20**, 323–340 (2019).
14. Krause, J. et al. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894–897 (2010).
15. Reich, D. et al. Genetic history of an archaic hominin group from Denisova cave in Siberia. *Nature* **468**, 1053–1060 (2010).
This article reports the discovery of Denisovans, a previously unknown lineage of archaic hominins that lived in southern Siberia at least 50,000 years ago. This was the first time that a representative of the *Homo* evolutionary tree was described from molecular data and in the absence of key macrofossil remains showing clear morphological characteristics.
16. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
This article presents the first high-quality genome from an archaic hominin and describes the first experimental procedure for DNA library preparation from single-stranded DNA templates. The approach outperforms competing technologies in sensitivity and complexity, and minimizes loss of authentic DNA molecules.
17. Rogers, R. L. & Slatkin, M. Excess of genomic defects in a woolly mammoth on Wrangel island. *PLoS Genet.* **13**, e1006601 (2017).
18. Lynch, V. J. et al. Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the Arctic. *Cell Rep.* **12**, 217–228 (2015).
19. Fry, E. et al. Functional architecture of deleterious genetic variants in the genome of a Wrangel island mammoth. *Genome Biol. Evol.* **12**, 48–58 (2020).
20. Lorenzen, E. D. et al. Species-specific responses of late quaternary megafauna to climate and humans. *Nature* **479**, 359–364 (2011).
21. Orlando, L. & Cooper, A. Using ancient DNA to understand evolutionary and ecological processes. *Annu. Rev. Ecol. Evol. Syst.* **45**, 573–598 (2014).
22. Willerslev, E. et al. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **506**, 47–51 (2014).
23. Fortes, G. G. et al. Ancient DNA reveals differences in behaviour and sociality between brown bears and extinct cave bears. *Mol. Ecol.* **25**, 4907–4918 (2016).
24. Gretzinger, J. et al. Large-scale mitogenomic analysis of the phylogeography of the Late Pleistocene cave bear. *Sci. Rep.* **9**, 10700 (2019).
25. Key, F. M. et al. Emergence of human-adapted *Salmonella enterica* is linked to the Neolithization process. *Nat. Ecol. Evol.* **4**, 324–333 (2020).
26. Warinner, C. et al. Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344 (2014).
This article reports the first metagenomic and paleoproteomic analysis of ancient dental plaque and demonstrates the preservation of oral microbial signatures, diet content and inflammation markers.
27. Warinner, C. et al. A robust framework for microbial archaeology. *Annu. Rev. Genomics Hum. Genet.* **18**, 321–356 (2017).
28. Gokhman, D. et al. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* **344**, 523–527 (2014).
29. Pedersen, J. S. et al. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* **24**, 454–466 (2014).
This article is the first report of ancient epigenomes, leveraging post-mortem DNA degradation signatures to statistically infer DNA methylation and nucleosome positioning.

30. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
31. Poinar, H. N. et al. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**, 392–394 (2006).
32. Green, R. E. et al. Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330–336 (2006).
33. Orlando, L., Gilbert, M. T. P. & Willerslev, E. Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* **16**, 395–408 (2015).
34. Giguet-Covex, C. et al. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat. Commun.* **5**, 3211 (2014).
35. Avila-Arcos, M. C. et al. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci. Rep.* **1**, 74 (2011).
36. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004 (2010).
37. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
This study presents the first application of in-solution target enrichment at the genome scale and describes sequence for a full chromosome of an approximately 40,000-year-old anatomically modern human from China.
38. Carpenter, M. L. et al. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am. J. Hum. Genet.* **93**, 852–864 (2013).
39. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
This study reports a new DNA extraction method from ancient osseous remains that is tailored to the ultrashort and extensively damaged nature of aDNA molecules. This methodology allowed the retrieval of full mitochondrial genome sequences from 300,000-year-old cave bear specimens preserved in Atapuerca, Spain.
40. Korlević, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *BioTechniques* **59**, 87–93 (2015).
41. Gamba, C. et al. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol. Ecol. Resour.* **16**, 459–469 (2016).
42. Boessenkool, S. et al. Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Mol. Ecol. Resour.* **17**, 742–751 (2017).
43. Gansauge, M.-T. & Meyer, M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.* **24**, 1543–1549 (2014).
44. Adler, C. J. et al. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and industrial revolutions. *Nat. Genet.* **45**, 450–455, 455e1 (2013).
45. Wagner, S. et al. High-throughput DNA sequencing of ancient wood. *Mol. Ecol.* **27**, 1138–1154 (2018).
46. Lendvay, B. et al. Improved recovery of ancient DNA from subfossil wood—application to the world's oldest Late Glacial pine forest. *New Phytol.* **217**, 1737–1748 (2018).
47. Der Sarkissian, C. et al. Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past. *Mol. Ecol. Resour.* **17**, 835–853 (2017).
48. Pedersen, M. W. et al. Postglacial viability and colonization in North America's ice-free corridor. *Nature* **537**, 45–49 (2016).
49. Slon, V. et al. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* **356**, 605–608 (2017).
50. Teasdale, M. D. et al. Paging through history: parchment as a reservoir of ancient DNA for next generation sequencing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130379 (2015).
51. O'Sullivan, N. J. et al. A whole mitochondria analysis of the Tyrolean Iceman's leather provides insights into the animal sources of Copper Age clothing. *Sci. Rep.* **6**, 31279 (2016).
52. Gamba, C. et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
This study reports the first matched empirical evidence establishing better DNA preservation rates in petrosal bones, which paved the way for future studies at the population scale, and also presents complete genomes of Mesolithic hunter–gatherers and Neolithic farmers in the Caucasus, supporting direct early contact between both groups.
53. Damgaard, P. B. et al. Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* **5**, 11184 (2015).
54. Krause, J. et al. A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* **20**, 231–236 (2010).
This article establishes important criteria for aDNA data authentication, including the presence of post-mortem cytosine deamination signatures and DNA fragmentation through depurination, and reports a complete mitochondrial genome from an approximately 32,000-year-old anatomically modern human from Kostenki, Russia.
55. Skoglund, P. et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234 (2014).
56. Renaud, G., Schubert, M., Sawyer, S. & Orlando, L. Authentication and assessment of contamination in ancient DNA. *Methods Mol. Biol.* **1963**, 163–194 (2019).
57. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
58. Leonardi, M. et al. Evolutionary patterns and processes: lessons from ancient DNA. *Syst. Biol.* **66**, e1–e29 (2017).
59. Pääbo, S., Gifford, J. A. & Wilson, A. C. Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Res.* **16**, 9775–9787 (1988).
60. Miller, W. et al. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**, 387–390 (2008).
61. Gilbert, M. T. P. et al. Resistance of degraded hair shafts to contaminant DNA. *Forensic Sci. Int.* **156**, 208–212 (2006).
62. Seguin-Orlando, A. et al. Pros and cons of methylation-based enrichment methods for ancient DNA. *Sci. Rep.* **5**, 11826 (2015).
63. Hagelberg, E., Sykes, B. & Hedges, R. Ancient bone DNA amplified. *Nature* **342**, 485 (1989).
64. Sirak, K. et al. Human auditory ossicles as an alternative optimal source of ancient DNA. *Genome Res.* **30**, 427–436 (2020).
65. Bos, K. I. et al. A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478**, 506–510 (2011).
This article reports the first complete genome of an ancient bacterial pathogen, *Y. pestis*, from osseous human remains of individuals who died from the Black Death in 1347–1348.
66. Margaryan, A. et al. Ancient pathogen DNA in human teeth and petrous bones. *Ecol. Evol.* **8**, 3534–3542 (2018).
67. Brown, T. A., Allaby, R. G., Brown, K. A., O'Donoghue, K. & Sallares, R. DNA in wheat seeds from European archaeological sites. *Experientia* **50**, 571–575 (1994).
68. Suyama, Y. et al. DNA sequence from a fossil pollen of *Abies* spp. from Pleistocene peat. *Genes. Genet. Syst.* **71**, 145–149 (1996).
69. Jaenicke-Després, V. et al. Early allelic selection in maize as revealed by ancient DNA. *Science* **302**, 1206–1208 (2003).
70. Vallebueno-Estrada, M. et al. The earliest maize from San Marcos Tehuacán is a partial domesticate with genomic evidence of inbreeding. *Proc. Natl Acad. Sci. USA* **113**, 14151–14156 (2016).
71. Ramos-Madrigal, J. et al. Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. *Curr. Biol.* **26**, 3195–3201 (2016).
72. Ramos-Madrigal, J. et al. Palaeogenomic insights into the origins of French grapevine diversity. *Nat. Plants* **5**, 595–603 (2019).
73. Weiß, C. L. et al. Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *R. Soc. Open Sci.* **3**, 160239 (2016).
74. Goloubinoff, P., Pääbo, S. & Wilson, A. C. Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proc. Natl Acad. Sci. USA* **90**, 1997–2001 (1993).
75. Mascher, M. et al. Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat. Genet.* **51**, 1076–1081 (2019).
76. Fordyce, S. L. et al. Deep sequencing of RNA from ancient maize kernels. *PLoS ONE* **8**, e50961 (2013).
77. Smith, O. et al. Small RNA activity in archeological barley shows novel germination inhibition in response to environment. *Mol. Biol. Evol.* **34**, 2555–2562 (2017).
78. Goldstein, P. Z. & Desalle, R. Calibrating phyfDabnlogenetic species formation in a threatened insect using DNA from historical specimens. *Mol. Ecol.* **12**, 1993–1998 (2003).
79. Rawlence, N. J., Wood, J. R., Armstrong, K. N. & Cooper, A. DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proc. Biol. Sci.* **276**, 3395–3402 (2009).
80. Oskam, C. L. et al. Fossil avian eggshell preserves ancient DNA. *Proc. Biol. Sci.* **277**, 1991–2000 (2010).
81. Bro-Jørgensen, M. H. et al. Ancient DNA analysis of Scandinavian medieval drinking horns and the horn of the last aurochs bull. *J. Archaeol. Sci.* **99**, 47–54 (2018).
82. Foley, B. P., Hansson, M. C., Kourkoumelis, D. P. & Theodoulou, T. A. Aspects of ancient Greek trade re-evaluated with amphora DNA evidence. *J. Archaeol. Sci.* **39**, 389–398 (2012).
83. Kashuba, N. et al. Ancient DNA from mastics solidifies connection between material culture and genetics of mesolithic hunter-gatherers in Scandinavia. *Commun. Biol.* **2**, 185 (2019).
84. Jensen, T. Z. T. et al. A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nat. Commun.* **10**, 5520 (2019).
85. Tito, R. Y. et al. Phylotyping and functional analysis of two ancient human microbiomes. *PLoS ONE* **3**, e3703 (2008).
86. Hagan, R. W. et al. Comparison of extraction methods for recovering ancient microbial DNA from paleofeces. *Am. J. Phys. Anthropol.* **171**, 275–284 (2020).
87. Søe, M. J. et al. Ancient DNA from latrines in Northern Europe and the Middle East (500 BC–1700 AD) reveals past parasites and diet. *PLoS ONE* **13**, e0195481 (2018).
88. Poinar, H. N. et al. Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* **281**, 402–406 (1998).
This article reports the first application of next-generation DNA sequencing to ancient specimens and describes the metagenomic nature of paleontological remains together with megabase-scale data from the woolly mammoth genome.
89. Bon, C. et al. Coprolites as a source of information on the genome and diet of the cave hyena. *Proc. Biol. Sci.* **279**, 2825–2830 (2012).
90. Willerslev, E. et al. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**, 791–795 (2003).
This article reports the first analysis of environmental DNA prior to the advent of next-generation DNA sequencing and establishes the long-term DNA persistence of paleocommunities within sediments.
91. Willerslev, E. et al. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* **317**, 111–114 (2007).
92. Coolen, M. J. & Overmann, J. Analysis of subfossil molecular remains of purple sulfur bacteria in a lake sediment. *Appl. Environ. Microbiol.* **64**, 4513–4521 (1998).
93. Bardill, J. et al. Advancing the ethics of paleogenomics. *Science* **360**, 384–385 (2018).
94. Prendergast, M. E. & Sawchuk, E. Boots on the ground in Africa's ancient DNA 'revolution': archaeological perspectives on ethics and best practices. *Antiquity* **92**, 803–815 (2018).
95. Alberti, F. et al. Optimized DNA sampling of ancient bones using computed tomography scans. *Mol. Ecol. Resour.* **18**, 1196–1208 (2018).
96. Sirak, K. A. et al. A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *BioTechniques* **62**, 283–289 (2017).
97. Harney, É. et al. A minimally destructive protocol for DNA extraction from ancient teeth. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.19.256412 (2020).
98. Patzold, F., Zilli, A. & Hundsdoerfer, A. K. Advantages of an easy-to-use DNA extraction method for minimal-destructive analysis of collection specimens. *PLoS ONE* **15**, e0235222 (2020).
99. Sugita, N. et al. Non-destructive DNA extraction from herbarium specimens: a method particularly suitable for plants with small and fragile leaves. *J. Plant Res.* **133**, 133–141 (2020).

100. Austin, R. M., Sholts, S. B., Williams, L., Kistler, L. & Hofman, C. A. Opinion: to curate the molecular past, museums need a carefully considered set of best practices. *Proc. Natl Acad. Sci. USA* **116**, 1471–1474 (2019).
101. Pálsdóttir, A. H., Bläuer, A., Rannamäe, E., Boessenkool, S. & Hallsson, J. H. Not a limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains. *R. Soc. Open Sci.* **6**, 191059 (2019).
102. Claw, K. G. et al. A framework for enhancing ethical genomic research with Indigenous communities. *Nat. Commun.* **9**, 2957 (2018).
103. Wagner, J. K. et al. Fostering responsible research on ancient DNA. *Am. J. Hum. Genet.* **107**, 183–195 (2020).
104. Eisenmann, S. et al. Reconciling material cultures in archaeology with genetic data: the nomenclature of clusters emerging from archaeogenomic analysis. *Sci. Rep.* **8**, 13003 (2018).
105. Frieman, C. J. & Hofmann, D. Present pasts in the archaeology of genetics, identity, and migration in Europe: a critical essay. *World Archaeol.* **51**, 528–545 (2019).
106. Furholt, M. De-contaminating the aDNA — archaeology dialogue on mobility and migration: discussing the culture-historical legacy. *CSA* **27**, 53–68 (2019).
107. Hakenbeck, S. E. Genetics, archaeology and the far right: an unholy Trinity. *World Archaeol.* **51**, 517–527 (2019).
108. Fulton, T. L. & Shapiro, B. Setting up an ancient DNA laboratory. *Methods Mol. Biol.* **1963**, 1–13 (2019).
109. Dabney, J., Meyer, M. & Pääbo, S. Ancient DNA damage. *Cold Spring Harb. Perspect. Biol.* **5**, a012567 (2013).
110. Höss, M. & Pääbo, S. DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Res.* **21**, 3913–3914 (1993).
111. Green, R. E. et al. The Neandertal genome and ancient DNA authenticity. *EMBO J.* **28**, 2494–2502 (2009).
112. Korlević, P., Talamo, S. & Meyer, M. A combined method for DNA analysis and radiocarbon dating from a single sample. *Sci. Rep.* **8**, 4127 (2018).
113. Fagernäs, Z. et al. A unified protocol for simultaneous extraction of DNA and proteins from archaeological dental calculus. *J. Archaeol. Sci.* **118**, 105135 (2020).
114. Rohland, N. & Hofreiter, M. Ancient DNA extraction from bones and teeth. *Nat. Protoc.* **2**, 1756–1762 (2007).
115. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018).
116. Glocke, I. & Meyer, M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Res.* **27**, 1230–1237 (2017).
117. Wales, N., Andersen, K., Cappellini, E., Ávila-Arcos, M. C. & Gilbert, M. T. P. Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLoS ONE* **9**, e86827 (2014).
118. Wales, N. et al. in *Ancient DNA: Methods and Protocols* (eds. Shapiro, B. et al.) 45–55 (Springer, 2019).
119. Meyer, M. et al. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403–406 (2014).
120. Höss, M., Jaruga, P., Zastawny, T. H., Dizdaroglu, M. & Pääbo, S. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.* **24**, 1304–1307 (1996).
121. Briggs, A. W. et al. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
This article presents the first statistical model of post-mortem DNA degradation and its impact on nucleotide misincorporation patterns, which provides the basis of important criteria for data authentication.
122. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* **7**, e34131 (2012).
123. Ho, S. Y. W., Heupink, T. H., Rambaut, A. & Shapiro, B. Bayesian estimation of sequence damage in ancient DNA. *Mol. Biol. Evol.* **24**, 1416–1422 (2007).
124. Axelsson, E., Willerslev, E., Gilbert, M. T. P. & Nielsen, R. The effect of ancient DNA damage on inferences of demographic histories. *Mol. Biol. Evol.* **25**, 2181–2187 (2008).
125. Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
This article presents an experimental procedure based on enzymatic digestion that eliminates nucleotide misincorporations from ancient sequence data, which provided the basis for robust statistical inference, minimally impacted by spurious allelic variation, and has paved the way to aDNA methylation maps.
126. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015).
This study reports the first method for DNA library preparation that is compatible with full automation and also describes various approaches aimed at authenticating data while minimizing the impact of post-mortem DNA misincorporation on downstream analyses.
127. Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. Biological identifications through DNA barcodes. *Proc. Biol. Sci.* **270**, 313–321 (2003).
128. Taberlet, P. et al. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **35**, e14 (2007).
129. Hofreiter, M., Mead, J. I., Martin, P. & Poinar, H. N. Molecular caving. *Curr. Biol.* **13**, R693–R695 (2003).
130. Parducci, L. et al. Glacial survival of boreal trees in northern Scandinavia. *Science* **335**, 1083–1086 (2012).
131. Alsos, I. G. et al. The role of sea ice for vascular plant dispersal in the Arctic. *Biol. Lett.* **12**, 20160264 (2016).
132. Crump, S. E. et al. Arctic shrub colonization lagged peak postglacial warmth: molecular evidence in lake sediment from Arctic Canada. *Glob. Chang. Biol.* **25**, 4244–4256 (2019).
133. Ziesemer, K. A. et al. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Sci. Rep.* **5**, 16498 (2015).
134. Smith, O. et al. Archaeology. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. *Science* **347**, 998–1001 (2015).
135. Parducci, L. et al. Shotgun environmental DNA, pollen, and macrofossil analysis of lateglacial lake sediments from southern Sweden. *Front. Ecol. Evol.* **7**, 189 (2019).
136. Gaffney, V. et al. Multi-proxy characterisation of the Storegga tsunami and its impact on the early holocene landscapes of the southern North Sea. *Geosciences* **10**, 270 (2020).
137. Cribdon, B., Ware, R., Smith, O., Gaffney, V. & Allaby, R. G. PIA: more accurate taxonomic assignment of metagenomic data demonstrated on sedaDNA from the North Sea. *Front. Ecol. Evol.* **8**, 84 (2020).
138. Noonan, J. P. et al. Genomic sequencing of Pleistocene cave bears. *Science* **309**, 597–599 (2005).
139. Noonan, J. P. et al. Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**, 1113–1118 (2006).
140. Bentley, D. R. et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
141. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb. prot5448 (2010).
142. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
143. Gansauge, M.-T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).
144. Gansauge, M.-T., Aximu-Petri, A., Nagel, S. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* **15**, 2279–2300 (2020).
145. Gansauge, M.-T. et al. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017).
146. Harkins, K. M. et al. A novel NGS library preparation method to characterize native termini of fragmented DNA. *Nucleic Acids Res.* **48**, e47 (2020).
147. Carøe, C. et al. Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419 (2018).
148. Fages, A. et al. Tracking five millennia of horse management with extensive ancient genome time series. *Cell* **177**, 1419–1435.e31 (2019).
149. van der Valk, T., Vezzi, F., Ormestad, M., Dalén, L. & Guschanski, K. Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. *Mol. Ecol. Resour.* **20**, 1171–1181 (2019).
150. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques* **52**, 87–94 (2012).
151. Seguin-Orlando, A. et al. Amplification of TruSeq ancient DNA libraries with AccuPrime Pfx: consequences on nucleotide misincorporation and methylation patterns. *STAR: Sci. Technol. Archaeol. Res.* **1**, 1–9 (2015).
152. Meyer, M. et al. From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res.* **36**, e5 (2008).
153. Thompson, J. R., Marcelino, L. A. & Polz, M. F. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res.* **30**, 2083–2088 (2002).
154. Max Planck Institute for the Science of Human History. MPI-SHH archaeogenetics. *protocols.io* https://www.protocols.io/workspaces/mpishh-archaeogenetics (2020).
155. Burbano, H. A. et al. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* **328**, 723–725 (2010).
156. Devault, A. M. et al. Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. *N. Engl. J. Med.* **370**, 334–340 (2014).
157. Enk, J. M. et al. Ancient whole genome enrichment using baits built from modern DNA. *Mol. Biol. Evol.* **31**, 1292–1294 (2014).
158. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
159. Cruz-Dávalos, D. I. et al. Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol. Ecol. Resour.* **17**, 508–522 (2017).
160. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
161. Cruz-Dávalos, D. I. et al. In-solution Y-chromosome capture-enrichment on ancient DNA libraries. *BMC Genomics* **19**, 608 (2018).
162. Günther, T. & Nettelblad, C. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genet.* **15**, e1008302 (2019).
163. Lachance, J. & Tishkoff, S. A. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *Bioessays* **35**, 780–786 (2013).
164. Lang, P. L. M. et al. Hybridization ddRAD-sequencing for population genomics of nonmodel plants using highly degraded historical specimen DNA. *Mol. Ecol. Resour.* **20**, 1228–1247 (2020).
165. Suchan, T. et al. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS ONE* **11**, e0151651 (2016).
166. Hernandez-Rodriguez, J. et al. The impact of endogenous content, replicates and pooling on genome capture from faecal samples. *Mol. Ecol. Resour.* **18**, 319–333 (2018).
167. Smith, O. et al. Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process. *Sci. Rep.* **4**, 5559 (2014).
168. Keller, A. et al. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **3**, 698 (2012).
169. Mak, S. S. T. et al. Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *Gigascience* **6**, 1–13 (2017).
170. Martins, R. F., Kampmann, M.-L. & Förster, D. W. Sequencing library preparation from degraded samples for non-illumina sequencing platforms. *Methods Mol. Biol.* **1963**, 85–92 (2019).
171. Kircher, M., Stenzel, U. & Kelso, J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol.* **10**, R83 (2009).

172. Renaud, G., Kircher, M., Stenzel, U. & Kelso, J. freelbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29**, 1208–1209 (2013).
173. Costello, M. et al. Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics* **19**, 332 (2018).
174. Dolle, D. et al. CASCADE: a custom-made archiving system for the conservation of ancient DNA experimental data. *Front. Ecol. Evol.* **8**, 185 (2020).
175. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).
176. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
177. Schuenemann, V. J. et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* **341**, 179–183 (2013).
178. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
179. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
180. Renaud, G., Hanghøj, K., Willerslev, E. & Orlando, L. gargammel: a sequence simulator for ancient DNA. *Bioinformatics* **33**, 577–579 (2017).
181. Schubert, M. et al. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
182. Shapiro, B. & Hofreiter, M. A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. *Science* **343**, 1236573 (2014).
183. Cahill, J. A. et al. Genomic evidence of widespread admixture from polar bears into brown bears during the last ice age. *Mol. Biol. Evol.* **35**, 1120–1129 (2018).
184. Poullet, M. & Orlando, L. Assessing DNA sequence alignment methods for characterizing ancient genomes and methylomes. *Front. Ecol. Evolut.* **8**, 105 (2020).
185. Seguin-Orlando, A. et al. Ligation bias in Illumina next-generation DNA libraries: implications for sequencing ancient genomes. *PLoS ONE* **8**, e78575 (2013).
186. Garrison, E. et al. Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat. Biotechnol.* **36**, 875–879 (2018).
187. Schubert, M. et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082 (2014).
188. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
189. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
190. Rasmussen, M. et al. An aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011).
191. Korneliussen, T. S., Moltke, I., Albrechtsen, A. & Nielsen, R. Calculation of Tajima's *D* and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinforma.* **14**, 289 (2013).
192. Peyrègne, S. & Peter, B. M. AuthentiCT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination. *Genome Biol.* **21**, 246 (2020).
193. Meyer, M. et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504–507 (2016).
194. Skoglund, P. et al. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469 (2012).
195. Nakatsuka, N. et al. A Paleogenomic reconstruction of the deep population history of the Andes. *Cell* **181**, 1131–1145.e21 (2020).
196. Skoglund, P., Ersmark, E., Palkopoulou, E. & Dalén, L. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr. Biol.* **25**, 1515–1519 (2015).
197. Soraggi, S., Wiuf, C. & Albrechtsen, A. Powerful inference with the *D*-statistic on low-coverage whole-genome data. *G3* **8**, 551–566 (2018).
198. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).
199. Mittnik, A., Wang, C.-C., Svoboda, J. & Krause, J. A molecular approach to the sexing of the triple burial at the upper Paleolithic site of Dolní Věstonice. *PLoS ONE* **11**, e0163019 (2016).
200. Schubert, M. et al. Zonkey: a simple, accurate and sensitive pipeline to genetically identify equine F1-hybrids in archaeological assemblages. *J. Archaeol. Sci.* **78**, 147–157 (2017).
201. Lipatov, M., Sanjeev, K., Patro, R. & Veeramah, K. Maximum likelihood estimation of biological relatedness from low coverage sequencing data. Preprint at *bioRxiv* https://doi.org/10.1101/023374 (2015).
202. Kennett, D. J. et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
203. Monroy Kuhn, J. M., Jakobsson, M. & Günther, T. Estimating genetic kin relationships in prehistoric populations. *PLoS ONE* **13**, e0195491 (2018).
204. Ziesemer, K. A. et al. The efficacy of whole human genome capture on ancient dental calculus and dentin. *Am. J. Phys. Anthropol.* **168**, 496–509 (2019).
205. Mittnik, A. et al. Kinship-based social inequality in Bronze Age Europe. *Science* **366**, 731–734 (2019).
206. Olalde, I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018); erratum **555**, 543 (2018).
207. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
208. McVean, G. A genealogical interpretation of principal components analysis. *PLoS Genet.* **5**, e1000686 (2009).
209. Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
210. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
211. Engelhardt, B. E. & Stephens, M. Analysis of population structure: a unifying framework and novel methods based on sparse factor analysis. *PLoS Genet.* **6**, e1001117 (2010).
212. Skoglund, P., Sjödin, P., Skoglund, T., Lascoux, M. & Jakobsson, M. Investigating population history using temporal genetic differentiation. *Mol. Biol. Evol.* **31**, 2516–2527 (2014).
213. Malaspinas, A.-S. et al. bammds: a tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS). *Bioinformatics* **30**, 2962–2964 (2014).
214. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
215. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
216. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
217. Joseph, T. A. & Pe'er, I. Inference of population structure from time-series genotype data. *Am. J. Hum. Genet.* **105**, 317–333 (2019).
218. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).
219. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012). **This article describes the statistical methodology forming the core of most subsequent aDNA studies and aims at investigating population structure and admixture patterns**.
220. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of native Americans. *Nature* **505**, 87–91 (2014).
221. Reich, D. et al. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
222. Cassidy, L. M. et al. A dynastic elite in monumental Neolithic society. *Nature* **582**, 384–388 (2020).
223. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
224. Antonio, M. L. et al. Ancient Rome: a genetic crossroads of Europe and the Mediterranean. *Science* **366**, 708–714 (2019).
225. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic studies. *Cell* **177**, 26–31 (2019).
226. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
227. Scheib, C. L. et al. East Anglian early Neolithic monument burial linked to contemporary Megaliths. *Ann. Hum. Biol.* **46**, 145–149 (2019).
228. Ringbauer, H., Novembre, J. & Steinrücken, M. Detecting runs of homozygosity from low-coverage ancient DNA. Preprint at *bioRxiv* https://doi.org/10.1101/2020.05.31.126912 (2020).
229. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
230. Burger, J. et al. Low prevalence of lactase persistence in Bronze Age Europe indicates ongoing strong selection over the last 3,000 years. *Curr. Biol.* **30**, 1–9 (2020).
231. Maixner, F. et al. The 5,300-year-old *Helicobacter pylori* genome of the Iceman. *Science* **351**, 162–165 (2016).
232. Lugli, G. A. et al. Ancient bacteria of the Ötzi's microbiome: a genomic tale from the Copper Age. *Microbiome* **5**, 5 (2017).
233. Borry, M. et al. CoproID predicts the source of coprolites and paleofeces using microbiome composition and host DNA content. *PeerJ* **8**, e9001 (2020).
234. Rifkin, R. F. et al. Multi-proxy analyses of a mid-15th century Middle Iron Age Bantu-speaker palaeo-faecal specimen elucidates the configuration of the 'ancestral' sub-Saharan African intestinal microbiome. *Microbiome* **8**, 62 (2020).
235. Tett, A. et al. The *Prevotella copri* complex comprises four distinct clades underrepresented in westernized populations. *Cell Host Microbe* **26**, 666–679.e7 (2019).
236. Vågene, Å. J. et al. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528 (2018).
237. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
238. Velsko, I. M., Frantz, L. A. F., Herbig, A., Larson, G. & Warinner, C. Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research. *mSystems* **3**, e00080-18 (2018).
239. Eisenhofer, R. & Weyrich, L. S. Assessing alignment-based taxonomic classification of ancient microbial DNA. *PeerJ* **7**, e6594 (2019).
240. Segata, N. et al. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811–814 (2012).
241. Hübler, R. et al. HOPS: automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biol.* **20**, 280 (2019).
242. Louvel, G., Der Sarkissian, C., Hanghøj, K. & Orlando, L. metaBIT, an integrative and automated metagenomic pipeline for analysing microbial profiles from high-throughput sequencing shotgun data. *Mol. Ecol. Resour.* **16**, 1415–1427 (2016).
243. Knights, D. et al. Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* **8**, 761–763 (2011).
244. Llamas, B. et al. High-resolution analysis of cytosine methylation in ancient DNA. *PLoS ONE* **7**, e30226 (2012).
245. Smith, R. W. A., Monroe, C. & Bolnick, D. A. Detection of cytosine methylation in ancient DNA from five Native American populations using bisulfite sequencing. *PLoS ONE* **10**, e0125344 (2015).
246. Wagner, S., Plomion, C. & Orlando, L. Uncovering signatures of DNA methylation in ancient plant remains from patterns of post-mortem DNA damage. *Front. Ecol. Evol.* **8**, 11 (2020).
247. Hanghøj, K. et al. Fast, accurate and automatic ancient nucleosome and methylation maps with epiPALEOMIX. *Mol. Biol. Evol.* **33**, 3284–3298 (2016).
248. Hanghøj, K., Renaud, G., Albrechtsen, A. & Orlando, L. DamMet: ancient methylome mapping accounting for errors, true variants, and post-mortem DNA damage. *Gigascience* **8**, giz025 (2019).
249. Pääbo, S. et al. Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38**, 645–679 (2004).
250. Racimo, F., Marnetto, D. & Huerta-Sánchez, E. Signatures of archaic adaptive introgression in present-day human populations. *Mol. Biol. Evol.* **34**, 296–317 (2017).
251. de Barros Damgaard, P. et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
252. Damgaard, PdeB. et al. 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018); author correction **563**, E16 (2018).

253. Narasimhan, V. M. et al. The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
254. Seguin-Orlando, A. et al. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
255. Sikora, M. et al. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662 (2017).
256. Fu, Q. et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
257. Yang, M. A. et al. 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. *Curr. Biol.* **27**, 3202–3208.e9 (2017).
258. Olalde, I. & Posth, C. Latest trends in archaeogenetic research of west Eurasians. *Curr. Opin. Genet. Dev.* **62**, 36–43 (2020).
259. Yang, M. A. et al. Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* **369**, 282–288 (2020).
260. Lipson, M. et al. Population turnover in remote Oceania shortly after initial settlement. *Curr. Biol.* **28**, 1157–1165.e7 (2018).
261. Lipson, M. et al. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* **361**, 92–95 (2018).
262. McColl, H. et al. The prehistoric peopling of Southeast Asia. *Science* **361**, 88–92 (2018).
263. Lazaridis, I. et al. Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
264. Wang, K. et al. Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa. *Sci. Adv.* **6**, eaaz0183 (2020).
265. Sánchez-Quinto, F. et al. Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr. Biol.* **22**, 1494–1499 (2012).
266. Lipson, M. et al. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).
267. Jeong, C. et al. The genetic history of admixture across inner Eurasia. *Nat. Ecol. Evol.* **3**, 966–976 (2019).
268. Moreno-Mayar, J. V. et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553**, 203–207 (2018).
269. Moreno-Mayar, J. V. et al. Early human dispersals within the Americas. *Science* **362**, eaav2621 (2018).
270. Flegontov, P. et al. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* **570**, 236–240 (2019).
271. Nieves-Colón, M. A. et al. Ancient DNA reconstructs the genetic legacies of precontact Puerto Rico communities. *Mol. Biol. Evol.* **37**, 611–626 (2020).
272. Nägele, K. et al. Genomic insights into the early peopling of the Caribbean. *Science* **369**, 456–460 (2020).
273. Schroeder, H. et al. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proc. Natl Acad. Sci. USA* **112**, 3669–3673 (2015).
274. Barquera, R. et al. Origin and health status of first-generation Africans from early Colonial Mexico. *Curr. Biol.* **30**, 2078–2091.e11 (2020).
275. Larson, G. et al. Current perspectives and the future of domestication studies. *Proc. Natl Acad. Sci. USA* **111**, 6139–6146 (2014).
276. Larson, G. et al. Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in island Southeast Asia and Oceania. *Proc. Natl Acad. Sci. USA* **104**, 4834–4839 (2007).
277. Kistler, L. et al. Ancient plant genomics in archaeology, herbaria, and the environment. *Annu. Rev. Plant. Biol.* **71**, 605–629 (2020).
278. Ní Leathlobhair, M. et al. The evolutionary history of dogs in the Americas. *Science* **361**, 81–85 (2018).
279. Frantz, L. A. F. et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* **352**, 1228–1231 (2016).
280. Sinding, M.-H. S. et al. Arctic-adapted dogs emerged at the Pleistocene–Holocene transition. *Science* **368**, 1495–1499 (2020).
281. Gaunitz, C. et al. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* **360**, 111–114 (2018).
282. Frantz, L. A. F. et al. Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. *Proc. Natl Acad. Sci. USA* **116**, 17231–17238 (2019).
283. Gutaker, R. M. et al. The origins and adaptation of European potatoes reconstructed from historical genomes. *Nat. Ecol. Evol.* **3**, 1093–1101 (2019).
284. Ameen, C. et al. Specialized sledge dogs accompanied Inuit dispersal across the North American Arctic. *Proc. R. Soc. Lond. B Biol. Sci.* **286**, 20191929 (2019).
285. da Fonseca, R. R. et al. The origin and evolution of maize in the Southwestern United States. *Nat. Plants* **1**, 14003 (2015).
286. Gutaker, R. M. et al. Flax latitudinal adaptation at LuTFL1 altered architecture and promoted fiber production. *Sci. Rep.* **9**, 976 (2019).
287. Park, S. D. E. et al. Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biol.* **16**, 234 (2015).
288. Verdugo, M. P. et al. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science* **365**, 173–176 (2019).
289. Daly, K. G. et al. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* **361**, 85–88 (2018).
290. Kistler, L. et al. Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. *Science* **362**, 1309–1313 (2018).
291. Allaby, R. G., Ware, R. L. & Kistler, L. A re-evaluation of the domestication bottleneck from archaeogenomic evidence. *Evol. Appl.* **12**, 29–37 (2019).
292. Scott, M. F. et al. A 3,000-year-old Egyptian emmer wheat genome reveals dispersal and domestication history. *Nat. Plants* **5**, 1120–1128 (2019).
293. Wales, N. et al. Ancient DNA reveals the timing and persistence of organellar genetic bottlenecks over 3,000 years of sunflower domestication and improvement. *Evol. Appl.* **12**, 38–53 (2018).
294. Smith, O. et al. A domestication history of dynamic adaptation and genomic deterioration in sorghum. *Nat. Plants* **5**, 369–379 (2019).
295. Swarts, K. et al. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science* **357**, 512–515 (2017).
296. Girdland Flink, L. et al. Establishing the validity of domestication genes using DNA from ancient chickens. *Proc. Natl Acad. Sci. USA* **111**, 6184–6189 (2014).
297. Loog, L. et al. Inferring allele frequency trajectories from ancient DNA indicates that selection on a chicken gene coincided with changes in medieval husbandry practices. *Mol. Biol. Evol.* **34**, 1981–1990 (2017).
298. Librado, P. et al. Ancient genomic changes associated with domestication of the horse. *Science* **356**, 442–445 (2017).
299. Salo, W. L., Aufderheide, A. C., Buikstra, J. & Holcomb, T. A. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc. Natl Acad. Sci. USA* **91**, 2091–2094 (1994).
300. Taubenberger, J. K. et al. Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).
301. Wilbur, A. K. et al. Deficiencies and challenges in the study of ancient tuberculosis DNA. *J. Archaeol. Sci.* **36**, 1990–1997 (2009).
302. Wagner, D. M. et al. *Yersinia pestis* and the plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect. Dis.* **14**, 319–326 (2014).
303. Spyrou, M. A. et al. Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. *Nat. Commun.* **10**, 4470 (2019).
304. Keller, M. et al. Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the first Pandemic (541–750). *Proc. Natl Acad. Sci. USA* **116**, 12363–12372 (2019).
305. Susat, J. et al. *Yersinia pestis* strains from Latvia show depletion of the pla virulence gene at the end of the second plague pandemic. *Sci. Rep.* **10**, 14628 (2020).
306. Bos, K. I. et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
307. Sabin, S. et al. A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex. *Genome Biol.* **21**, 201 (2020).
308. Shillito, L.-M., Blong, J. C., Green, E. J. & van Asperen, E. N. The what, how and why of archaeological coprolite analysis. *Earth Sci. Rev.* **207**, 103196 (2020).
309. Warinner, C., Speller, C. & Collins, M. J. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130376 (2015).
310. Velsko, I. M. et al. Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome* **7**, 102 (2019).
311. Farrer, A. G. et al. Biological and cultural drivers of oral microbiota in Medieval and Post-Medieval London, UK. Preprint at *bioRxiv* https://doi.org/10.1101/343889 (2018).
312. Ho, S. Y. W. & Shapiro, B. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Resour.* **11**, 423–434 (2011).
313. Allentoft, M. E. et al. Extinct New Zealand megafauna were not in decline before human colonization. *Proc. Natl Acad. Sci. USA* **111**, 4922–4927 (2014).
314. Thomas, J. E. et al. Demographic reconstruction from ancient DNA supports rapid extinction of the great auk. *eLife* **8**, e47509 (2019).
315. Dussex, N. et al. Complete genomes of two extinct New Zealand passerines show responses to climate fluctuations but no evidence for genomic erosion prior to extinction. *Biol. Lett.* **15**, 20190491 (2019).
316. Li, H. & Durbin, R. Inference of human population history from whole genome sequence of a single individual. *Nature* **475**, 493–496 (2011).
317. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
318. Murray, G. G. R. et al. Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science* **358**, 951–954 (2017).
319. Díez-Del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P. & Dalén, L. Quantifying temporal genomic erosion in endangered species. *Trends Ecol. Evol.* **33**, 176–185 (2018).
320. Der Sarkissian, C. et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr. Biol.* **25**, 2577–2583 (2015).
321. Norén, K., Godoy, E., Dalén, L., Meijer, T. & Angerbjörn, A. Inbreeding depression in a critically endangered carnivore. *Mol. Ecol.* **25**, 3309–3318 (2016).
322. Pedersen, M. W. et al. Ancient and modern environmental DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130383 (2015).
323. Haile, J. et al. Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? *Mol. Biol. Evol.* **24**, 982–989 (2007).
324. Binladen, J. et al. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* **2**, e197 (2007).
325. Murchie, T. J. et al. Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. *Quat. Res.* https://doi.org/10.1017/qua.2020.59 (2020).
326. Haile, J. et al. Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proc. Natl Acad. Sci. USA* **106**, 22352–22357 (2009).
327. Seersholm, F. V. et al. Rapid range shifts and megafaunal extinctions associated with late Pleistocene climate change. *Nat. Commun.* **11**, 2770 (2020).
328. Key, F. M., Posth, C., Krause, J., Herbig, A. & Bos, K. I. Mining metagenomic data sets for ancient DNA: recommended protocols for authentication. *Trends Genet.* **33**, 508–520 (2017).
329. Günther, T. & Jakobsson, M. in *Handbook of Statistical Genomics* (eds Balding, D., Moltke, I. & Marioni, J.) 295–340 (Wiley, 2019).
330. Skourtanioti, E. et al. Genomic history of Neolithic to bronze age Anatolia, Northern Levant, and Southern Caucasus. *Cell* **181**, 1158–1175.e28 (2020).
331. Burmeister, S. Archaeological research on migration as a multidisciplinary challenge. *medieval worlds* **4**, 42–64 (2016).
332. Knipper, C. et al. Female exogamy and gene pool diversification at the transition from the final Neolithic to the early bronze age in central Europe. *Proc. Natl Acad. Sci. USA* **114**, 10083–10088 (2017).
333. Zarrillo, S. et al. The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nat. Ecol. Evol.* **2**, 1879–1888 (2018).
334. Hendy, J. et al. Proteomic evidence of dietary sources in ancient dental calculus. *Proc. Biol. Sci.* **285**, 20180977 (2018).
335. Roberts, P. et al. Calling all archaeologists: guidelines for terminology, methodology, data handling, and reporting when undertaking and reviewing stable isotope applications in archaeology. *Rapid Commun. Mass. Spectrom.* **32**, 361–372 (2018).

336. Radini, A. et al. Medieval women's early involvement in manuscript production suggested by lapis lazuli identification in dental calculus. *Sci. Adv.* **5**, eaau7126 (2019).
337. Ren, M. et al. The origins of cannabis smoking: chemical residue evidence from the first millennium BCE in the Pamirs. *Sci. Adv.* **5**, eaaw1391 (2019).
338. Reitsema, L. J. & Holder, S. Stable isotope analysis and the study of human stress, disease, and nutrition. *Bioarc. Int.* **2**, 63–74 (2018).
339. Maixner, F. et al. The Iceman's last meal consisted of fat, wild meat, and cereals. *Curr. Biol.* **28**, 2348–2355.e9 (2018).
340. Lejzerowicz, F. et al. Ancient DNA complements microfossil record in deep-sea subsurface sediments. *Biol. Lett.* **9**, 20130283 (2013).
341. van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. & Thermes, C. The third revolution in sequencing technology. *Trends Genet.* **34**, 666–681 (2018).
342. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
343. Li, H. L., Gee, P., Ishida, K. & Hotta, A. Efficient genomic correction methods in human iPS cells using CRISPR–Cas9 system. *Methods* **101**, 27–35 (2016).
344. Muchnik, S. K., Lorente-Galdos, B., Santpere, G. & Sestan, N. Modeling the evolution of human brain development using organoids. *Cell* **179**, 1250–1253 (2019).
345. Kanton, S. et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).
346. Racimo, F., Renaud, G. & Slatkin, M. Joint estimation of contamination, error and demography for nuclear DNA from ancient humans. *PLoS Genet.* **12**, e1005972 (2016).
347. Jun, G. et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **91**, 839–848 (2012).
348. Link, V. et al. ATLAS: analysis tools for low-depth and ancient samples. Preprint at *bioRxiv* https://doi.org/10.1101/105346 (2017).
349. Librado, P. & Orlando, L. Detecting signatures of positive selection along defined branches of a population tree using LSD. *Mol. Biol. Evol.* **35**, 1520–1535 (2018).
350. Refoyo-Martínez, A. et al. Identifying loci under positive selection in complex population histories. *Genome Res.* **29**, 1506–1520 (2019).
351. Renaud, G., Hanghøj, K., Korneliussen, T. S., Willerslev, E. & Orlando, L. Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. *Genetics* **212**, 587–614 (2019).
352. Brace, S. et al. Ancient genomes indicate population replacement in early neolithic britain. *Nat. Ecol. Evol.* **3**, 765–771 (2019).
353. Yates, J. A. F. et al. Community-curated and standardised metadata of published ancient metagenomic samples with AncientMetagenomeDir. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.02.279570 (2020).

Acknowledgements

The authors thank A. Hübner for assistance with figure 4c. L.O., P.S., P.W.S. and C.W. received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreements ERC-2015-CoG 681605-PEGASUS, ERC-2018-StG 852558-AGRICON, ERC-2015-StG 678901-FoodTransforms and ERC-2017-StG 804844-DAIRYCULTURES, respectively). L.O. was also supported by ANR (LifeChange) and the Simone et Cino Del Duca Foundation (HealthTimeTravel). P.S. was also supported by the Francis Crick Institute core funding (FC001595) from Cancer Research UK, the UK Medical Research Council and the Wellcome Trust, a Wellcome Trust Investigator award (217223/Z/19/Z) and the Vallee Foundation. C.W. also received funding from the Max Planck Society, the Deutsche Forschungsgemeinschaft (EXC 2051 #390713860) and the Siemens Foundation (Paleobiochemistry).

Author contributions

Introduction (L.O., A.C.S. and C.W.); Experimentation (L.O., R.A., C.D.S., P.W.S., A.C.S. and C.W.); Results (L.O., P.S., R.A., P.W.S., M.C.A.-A. and C.W.); Applications (L.O., R.A., P.W.S., C.D.S., M.C.A.-A., Q.F., J.K., E.W., A.C.S. and C.W.); Reproducibility and data deposition (L.O. and M.C.A.-A.); Limitations and optimizations (L.O.); Outlook (L.O., P.W.S., A.C.S. and C.W.); Overview of the Primer (L.O. and C.W.).

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Methods Primers thanks T. Günther, L. Matisoo-Smith, R. Pinhasi, N. Rawlence, A. Zink and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

RELATED LINKS

Ancient Human DNA uMap: https://umap.openstreetmap.fr/en/map/ancient-human-dna_41837#6/51.000/2.000
Bitbucket: https://bitbucket.org/product
European Nucleotide Archive (ENA): https://www.ebi.ac.uk/ena
GitHub: https://github.com
International Symposium for Biomolecular Archaeology (ISBA): https://isba9.sciencesconf.org
Max Planck Harvard Research Center for the Archaeoscience of the Ancient Mediterranean: https://www.archaeoscience.org/
Protocols.io: https://www.protocols.io
Reich laboratory: https://reich.hms.harvard.edu/datasets
Sequence Read Archive (SRA): https://www.ncbi.nlm.nih.gov/sra
Standards and Precautions and Advances in Ancient Metagenomics (SPAAM): https://github.com/SPAAM-workshop