

Review

Examining Natural History through the Lens of Palaeogenomics

Kieren J. Mitchell ^{1,2,*} and Nicolas J. Rawlence³

The many high-resolution tools that are uniquely applicable to specimens from the Quaternary period (the past ~2.5 Ma) provide an opportunity to cross-validate data and test hypotheses based on the morphology and distribution of fossils. Among these tools is palaeogenomics – the genome-scale sequencing of genetic material from ancient specimens – that can provide direct insight into ecology and evolution, potentially improving the accuracy of inferences about past ecological communities over longer timescales. Palaeogenomics has revealed instances of over- and underestimation of extinct diversity, detected cryptic faunal migration and turnover, allowed quantification of widespread sex biases and sexual dimorphism in the fossil record, revealed past hybridisation events and hybrid individuals, and has highlighted previously unrecognised routes of zoonotic disease transfer.

Quaternary Science and the Advent of Palaeogenomics

The morphology of fossilised remains constitutes our main source of data about ecosystem composition and function for >99% of life's history [1], making palaeontology a crucial and irreplaceable discipline for studying evolution. However, the fossil record is inherently limited in several respects that must be overcome by palaeontologists. For example, many extinct species are described from small numbers of specimens, often drawn from temporally or geographically disparate sources, and whose occurrence is affected by several **taphonomic biases** (see [Glossary](#)), limiting the ability to accurately determine crucial information on intra- versus inter-species morphological variation. Further, morphological variation may be obscured by convergent evolution, functional constraint, or a lack of differentiation among recently diverged taxa. In addition, directly establishing the age of specimens is often challenging, and the uncertainty associated with some dating methods is often wide relative to the pace of many biological processes. These factors complicate the process of temporally linking specimens and climate proxies, which often have coarse resolution, in turn making it difficult to identify cohesive ecological units. Finally, hypotheses about past ecological functions and drivers of ecosystem change must be based on comparisons to better-understood extant analogues, which in many cases simply do not exist. Nevertheless, palaeontologists have made admirable progress towards addressing these limitations.

The **Quaternary period** (2.58 Mya to present) represents a valuable natural laboratory with numerous advantages that allow hypotheses based on the morphology of biological remains preserved in the fossil record to be independently tested. For example, Quaternary scientists have access to more precise methods for establishing the age of specimens, climate proxies that are not available from earlier periods, and generally higher-resolution deposits that provide greater sample sizes. These different records can be cross-referenced to not only provide a more holistic understanding of past ecological changes but also to identify and quantify errors or biases in individual records. In addition, since the 1980s, the sequencing of DNA from organisms up to several hundreds of thousands of years old, so-called **ancient DNA** (aDNA), has become another tool that is routinely used by scientists studying Quaternary ecosystems.

Highlights

Under ideal conditions, it is possible to retrieve genomic data from the remains of organisms hundreds of thousands of years old. These 'palaeogenomic' data can be used to test hypotheses about past biological change with a level of resolution that is not possible using other methods.

Palaeogenomic studies have revealed underappreciated taxonomic diversity and intraspecific morphological variation in past ecological communities, as well as documenting processes such as hybridisation and migration that are challenging to infer from fossilised remains.

Although the temporal range of palaeogenomic studies is limited by the half-life of DNA, the deeper understanding of biases in the recent fossil record provided by cross-referencing with palaeogenomic data is likely to stimulate reevaluation of patterns observed in the fossil record over much longer time-periods.

¹Australian Centre for Ancient DNA (ACAD), School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia

²Australian Research Council (ARC) Centre of Excellence for Australian Biodiversity and Heritage (CABAH), School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia

³Otago Palaeogenetics Laboratory, Department of Zoology, University of Otago, Dunedin, New Zealand

*Correspondence: kieren.mitchell@adelaide.edu.au (K.J. Mitchell).



Following the successful sequencing of short fragments of the **mitochondrial genome** from a 100-year-old quagga (*Equus quagga*) specimen [2], the field of aDNA expanded rapidly. As technology advanced, and the field adopted rigorous methodological standards [3,4], aDNA studies expanded from short fragments of mitochondrial DNA to complete mitochondrial genome sequences from a diverse range of taxa [5,6]. Other studies went even further and sequenced portions of the **nuclear genome** [7–12]. However, these nuclear data generally only represented a tiny proportion of the total length of the genome (i.e., <1%). Over the past decade, continuing technological advances in **high-throughput DNA sequencing** have finally made it economically and technically feasible to obtain highly accurate complete nuclear genome sequences from ancient specimens, or at the very least data that represent a substantial proportion of the nuclear genome. This new discipline, **palaeogenomics**, promises to provide a rich new source of data for Quaternary scientists.

Although the palaeogenomic revolution began slightly earlier in the field of human evolutionary research [13], including the publication of the Neandertal genome sequence in 2010 [14], it was not long before similar methods were applied to other extinct taxa. Indeed, in 2013 the complete genome of a 560 000–780 000 year old horse (*Equus* spp.) was sequenced to 1.12× average **depth of coverage** [15], which to date represents the oldest specimen from which palaeogenomic data are available (Figure 1). Although the horse specimen was collected from **permafrost**, which provides ideal conditions for long-term DNA preservation, it is now clear that data of similar quality can be obtained from very old specimens from more temperate localities. For example, palaeogenomic data have recently been reported from a 277 000–457 000 year old cave bear (*Ursus* spp.) specimen excavated in the Southern Caucasus [16] and >120 000 year old specimens of the straight-tusked elephant (*Palaeoloxodon antiquus*) from Germany [17] (Figure 1). Palaeogenomic data have already provided insight into the process of domestication [18,19], the phylogenetic relationships and evolutionary adaptations of extinct species [20,21,93,94], the mechanisms underlying the loss of flight in palaeognathous birds [22], the convergence between carnivorous marsupials and placental mammals [23,24], and the origin of some of humanity's deadliest diseases [25]. However, in this review, we will examine how the palaeogenomic revolution can contribute new insights into past ecological communities and provide a new dimension to Quaternary science that can inform our interpretation of fossil data more broadly.

Accurately Characterising Taxonomic Diversity

Taxonomy is one of the cornerstones of ecology and evolutionary biology because the taxon to which an organism belongs reflects its identity and relatedness to other organisms. Taxonomy also allows the diversity of ecological communities to be quantified and compared. Consequently, accurate taxonomic designation is crucial for looking at past ecosystem changes. By necessity, the taxonomy of most extinct species is primarily based on levels of skeletal variation because other types of data – concerning plumage/pelage, behaviour, diet, or genetics – are usually not available. As a result, the process of assigning a fossil specimen to a taxon must be informed by comparison to intra- and inter-taxon skeletal variation in living organisms from related taxa, where taxonomic assignments can be verified using other types of information. However, it is likely that the morphological variation of some extant taxa is impoverished as a result of recent anthropogenic biodiversity loss. In addition, for extant taxa we generally only have detailed data spanning at most the past several hundred years, whereas a given taxon may have persisted and undergone morphological change over hundreds of millennia. These factors mean that basing fossil taxonomic assignment on thresholds of variation observed in extant taxa may lead to inaccurate measures of past taxonomic diversity. The application of palaeogenomic methods to Quaternary taxa therefore provides an opportunity to measure and verify levels of 'natural' morphological variation (i.e., before modern biodiversity loss) in taxa over timescales relevant to the fossil record.

Glossary

Ancient DNA (aDNA): poorly preserved, fragmented, and chemically altered DNA that is usually present only in trace amounts in an ancient biological or environmental sample.

Dental calculus: a calcified biofilm created by oral microbial communities, which may preserve microbial DNA for thousands of years.

Depth of coverage: the average number of times that the sequence of an individual position in the genome is independently represented in high-throughput DNA sequencing data.

Genetic load: an accumulation of harmful or deleterious genetic variants or mutations in a population or individual.

High-throughput DNA sequencing: technology that allows the nucleotide sequence of hundreds of millions to billions of DNA fragments to be obtained in parallel, dramatically increasing the speed and decreasing the cost of genome sequencing projects.

Hybridisation: interbreeding between individuals belonging to different species, which can result in fertile offspring and gene flow between distantly related taxa.

Megafauna: large animal species, often defined as those with adult body mass of >44 kg, but sometimes used simply to refer to the largest animal species in an ecosystem.

Mitochondrial genome: a comparatively small DNA molecule found within the mitochondria of eukaryotic cells. Many more copies of the mitochondrial genome are present in a given cell than for the nuclear genome, making it easier to sequence and assemble from ancient specimens.

Nuclear genome: the total DNA within the nucleus of eukaryotic cells, accounting for the majority of genetic material belonging to most multicellular organisms. The size of the nuclear genome often makes it challenging to sequence and assemble from ancient specimens with poor DNA preservation.

Ontogeny: the morphological changes that occur, especially to bones, from birth to adulthood (i.e., osteological maturity).

Palaeoenvironmental DNA: aDNA originating from disseminated genetic material within palaeoenvironmental samples, including but not limited to sediment, soil, calcite, dental calculus, and coprolites.

Even before accounting for morphological change in a taxon through time, two major sources of intraspecies morphological variation are **ontogeny** and **sexual dimorphism**. Ontogeny is relatively well studied in some non-avian dinosaur species and can be resolved by using histology and/or collecting sufficient specimens to reconstruct a continuous developmental series [26]. By contrast, sexual dimorphism frequently results in a multimodal distribution of morphological characters, which in some circumstances could be interpreted as representing two or more different taxa. This was the case in the extinct giant moa (*Dinornis* spp.) from New Zealand, where **polymerase chain reaction (PCR)** of a sex-linked gene sequence from ancient specimens revealed that three geographically widespread species described based on skeletal morphology instead represented sex-linked variation within two different island endemics [27,28]. Recent palaeogenomic studies have sought to distinguish the sex of individuals by adopting a 'read-dosage' approach, which involves looking at the ratio of sequenced reads mapping to the sex chromosomes versus autosomes [29,30]. Although originally applied to human remains, the read-dosage approach has the twin benefits of working on practically any taxon with heteromorphic sex chromosomes and requiring minimal input DNA, meaning that the method can successfully be applied up to the temporal limits of DNA preservation. For example, Gower *et al.* [31] and Pečnerová *et al.* [32] used read-dosage to determine the sex of hundreds of Quaternary specimens of bison (*Bison* spp.), brown bears (*Ursus arctos*), and mammoths (*Mammuthus* spp.), including many that were >50 000 years old, and detected around threefold as many males as females across all three taxa. Another palaeogenomic study looking explicitly at sexual dimorphism in extinct musk-oxen (*Bootherium bombifrons*, *Symbos cavifrons*) revealed a similar bias towards males in palaeontological collections, and also confirmed that size variation in specimens previously assigned to two different genera actually belonged to males and females of the same species [33]. An important implication of these results more broadly is that using ratios of differently sized specimens to diagnose sexual dimorphism versus different taxa in the absence of genetic data may not yield the expected results.

Other aDNA studies have revealed substantial intraspecies morphological variation through space and time. For example, whereas bison in North America today are represented by only a single species – the American bison (*Bison bison*) – remains from the Quaternary have been assigned to several extinct species, including the steppe bison (*B. priscus*), ancient bison (*B. antiquus*), and giant long-horned bison (*B. latifrons*). However, published mitochondrial data do not support these species-level divisions [34–36]. Similarly, in mammoths, published mitochondrial data appear to support the existence of three main Quaternary lineages in Eurasia and North America, but these lineages do not appear to conform to recognised species [37,38]. One plausible interpretation of these data is that mammoths and North American bison each represent only a single Quaternary species, which would imply that intraspecies skeletal diversity in these taxa over millennial timescales has been drastically underappreciated. The application of similar thresholds across fossil mammals more broadly would likely necessitate **synonymisation** of many species-level taxa, although it would be premature to do so based only on mitochondrial results. Indeed, in some cases palaeogenomic data have actually revealed that extinct Quaternary species are more distinct than was inferred from either morphological or mitochondrial data. For example, an extinct lineage of endemic North American horses was recently raised to its own monotypic genus, *Haringtonhippus* [39], and palaeogenomic data from the extinct dire wolf (*Canis dirus*) have revealed it was only a distant relative to the morphologically similar grey wolf, *Canis lupus* [40].

Palaeogenomic data (e.g., [16,41,42]) as well as genomic data from extant populations [43] are increasingly demonstrating that **hybridisation** between biological species has been common

Palaeogenomics: the subfield focused on aDNA sequencing and the analysis of genome-scale genetic data.

Palaeoproteomics: a field of research and a suite of methods focused on reconstructing the amino acid sequences of ancient protein molecules.

Permafrost: soil or sediment that remains frozen year-round, commonly found in high-latitude areas.

Polymerase chain reaction (PCR): a fundamental molecular biology technique that allows the user to produce millions of copies of a specific target DNA molecule through enzymatic amplification.

Punctuated equilibrium: the observation that some fossil lineages appear to exhibit long periods of morphological stasis, interrupted by brief periods of rapid change, followed by a return to stasis.

Quaternary period: the most recent period of Earth's geological history, encompassing the Pleistocene (2.58 million to 11.7 thousand years ago) and the Holocene (11.7 thousand years ago to present) epochs.

Sexual dimorphism: differences expressed between the sexes of a species of population, frequently manifesting as variation in size, pelage, or plumage.

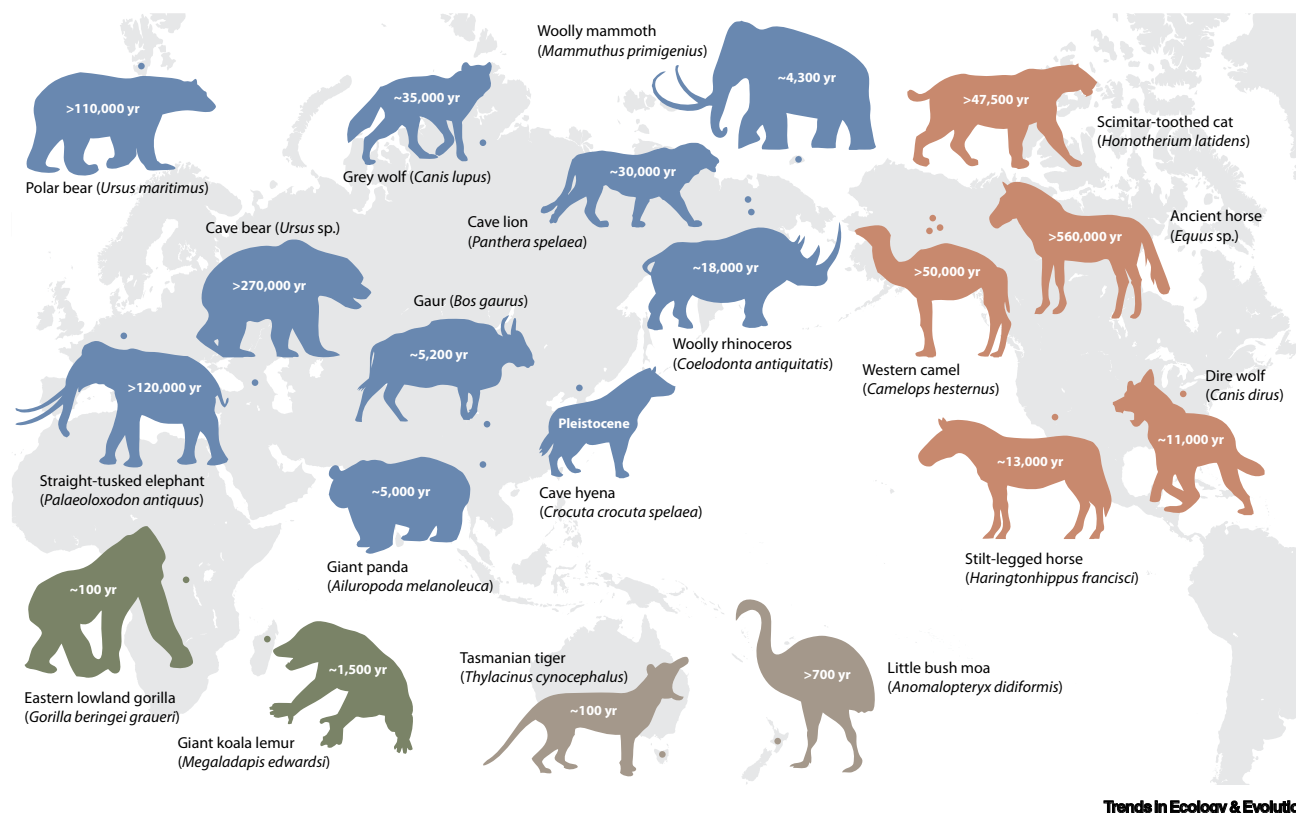
Shotgun sequencing: sequencing of a random – and theoretically representative – subset of the DNA molecules extracted from a sample, regardless of their origin or taxonomic identity.

Synonymisation: when evidence suggests that multiple names have been proposed for a single taxon, these names are synonymous, and all but one will usually be removed from common usage, and past usages may be revised/reinterpreted.

Taphonomic biases: biological and geological processes that occur between death, decay, and eventual fossilisation that result in over- or under-representation or even complete absence of specific skeletal elements, taxa, or demographic groups (e.g., different sexes or ages) in the fossil record.

Taxonomy: the scientific discipline of classifying life under the Linnaean hierarchy, notably including genus and species.

Zoonotic/zoonoses: pathogens – including viruses, bacteria, microbes, and multicellular parasites – that can spread from animal to animal or from animal to human.



Trends in Ecology & Evolution

Figure 1. Geographical Distribution, Taxonomic Identity, and Approximate Age of Some Key Animal Specimens from which Palaeogenomic Data Have Been Generated. Silhouettes not to scale; references are given in the text.

during the Quaternary (although not ubiquitous (e.g., [21,93]), and thus represents an alternative hypothesis for reconciling the observed conflicts between taxonomy, morphological variation, and mitochondrial phylogenies in Quaternary taxa. It is possible that in some cases gene flow followed by fixation of an introgressed mitochondrion or Y chromosome in descendants of the hybridisation event has obscured the 'true' phylogenetic signal that more closely matches the accepted taxonomy, as observed in hominins [44] and European bison (*Bison bonasus*) [45], as well as polar bears (*Ursus maritimus*) and brown bears [46–48]. Indeed, although the hypothesis remains to be tested using palaeogenomic data, it has already been proposed that many mammoth specimens may represent individuals with hybrid ancestry from both woolly mammoths (*Mammuthus primigenius*) and Columbian mammoths (*M. columbi*) [49]. This innocuous suggestion – although highly plausible based on evidence for gene flow between extinct straight-tusked elephants and mammoths [17,50] – carries an important implication for interpreting morphological variation: hybrid individuals may be sufficiently common in nature that we should expect to sample them in the fossil record. This is supported by the detection of a first-generation Neandertal (*Homo neanderthalensis*)–Denisovan hybrid among the small handful of archaic hominins from which palaeogenomic data have been generated to date [51]. Specimens or lineages with hybrid ancestry may have intermediate or aberrant morphology that could in some circumstances be difficult to distinguish from a novel taxon. In some cases, the morphological or physiological results of hybridisation may even increase the fitness of hybrid descendants, such as the adaptation for living at high altitudes that modern Tibetans inherited from Denisovans [52], resulting in lasting changes to a population or species. Lasting morphological changes driven by

hybridisation may occur sufficiently rapidly to appear instantaneous when viewed in the fossil record over long timescales, which in some instances may underlie observations of 'sudden jumps' in the evolution of taxa that otherwise appear to have undergone only gradual morphological change (or stasis), a phenomenon termed **punctuated equilibrium**.

Changes in Distribution and Demography on a Millennial Scale

It is well established that the Quaternary history of hominins – including that of our direct ancestors – has been hugely dynamic (e.g., [13,53]), with repeated waves of intercontinental migration and extinction, but an important finding of recent aDNA studies has been that many animal species have similarly complex histories. Indeed, the distribution of non-human species through time has been unexpectedly dynamic, and genetic data from Quaternary remains reveal widespread movement and turnover – including extirpation and replacement – that are undetectable based on either morphological data or the genetics of modern populations [34,37,54,95]. For example, analyses of the mitogenomes from grey wolves have suggested that all modern individuals descend from a population that migrated out of northwestern Eurasia and/or northeastern North America ~25 000 years ago [55], despite a continuous Holarctic fossil record spanning the past 300 000 years [56]. However, mitochondrial data represent only a single maternally inherited non-recombining locus, meaning that they do not always reveal the nuances of past population processes. Consequently, it is possible that local populations of grey wolves in Eurasia still derive substantial nuclear genomic ancestry from much older Quaternary populations that remained *in situ*, though available palaeogenomic data suggest this is unlikely to be the case [96]. Conversely, palaeogenomic data from a 5000 year old giant panda (*Ailuropoda melanoleuca*) specimen, from outside the current range of the species, revealed a genetic contribution and connection with modern populations that are not reflected in the mitochondrial data [57]. In any case, the use of powerful methods for investigating the speed and extent of past migration events and population turnover using palaeogenomic data [58] may help to better understand the process of dispersal inferred from the fossil record, and species-specific data may provide important long-term baseline data for evidence-based conservation management where records of past distribution are lacking [59].

Extinction events during the Quaternary – in particular the global extinction of terrestrial **megafauna** – have long been recognised as a powerful model for studying the drivers of population declines [60]. Palaeogenomic data add a new dimension by allowing the reconstruction of long-term changes in population size (i.e., over hundreds of millennia (e.g., [61,62]) and measurement of **genetic load**, which may contribute to lowered fitness, in the lead up to the extinction of a species (e.g., [63]). For example, palaeogenomic data from mammoths on Wrangel Island, where the species survived well into the Holocene, allowed quantification of genetic diversity loss [64] and the accumulation of genetic defects [65] that are likely implicated in their eventual extinction. A similar study of the woolly rhinoceros (*Coelodonta antiquitatis*) suggested that it was ultimately driven extinct by rapid warming during the Bølling–Allerød interstadial, and that the arrival of humans in Siberia had no appreciable demographic impact [62]. By contrast, palaeogenomic data from two species of extinct passerine birds from New Zealand revealed population fluctuations that correlated with climate and environmental change, but no signals of genomic erosion, suggesting that anthropogenic introduction of mammalian predators was the primary driver of their extinction [66]. Further studies of palaeogenomic data, especially from extinct species, will allow better characterisation of the population trajectories that may lead to either survival or extinction through major perturbations, and could allow more targeted conservation efforts, particularly with respect to attenuating genetic load (e.g., genetic rescue [67]). In addition, comparing palaeogenomic reconstructions of past population size to the observed frequency of fossil remains may allow better quantification of taphonomic biases and more accurate inference of extinction dates from the fossil record.

Environmental Palaeogenomic Records

A major limitation of the fossil record is that specimens – especially skeletal remains of large vertebrates – are relatively rare, and their age and distribution are influenced not only by taxon-specific biology and behaviour but also by taphonomic biases that determine preservation probability. Trace DNA extracted from environmental samples such as sediment, water, or faeces – environmental DNA – has the power to reveal the presence of taxa indirectly, and **palaeoenvironmental DNA** from ancient samples can therefore be used to determine past ecological community composition even in the absence of rare physical remains. Indeed, Slon *et al.* [68] obtained genetic data from a series of Eurasian cave sediments ranging in age from ~14 000 to >100 000 years old, which revealed the presence of several extinct megafaunal taxa (e.g., cave bear, woolly mammoth, woolly rhino) even from strata containing no diagnostic faunal remains, and suggested that occupation of Denisova Cave in the Altai Mountains alternated between Neandertals and Denisovans. Seersholm *et al.* [69] used palaeoenvironmental DNA data from sediment and 'bulk-bone' – pooled non-diagnostic bone fragments [70] – from Hall's Cave in North America to measure species-level community structure and responses to local climate and environmental change, and importantly demonstrated the existence of past Quaternary ecosystems that have no modern analogue. Finally, Murchie *et al.* [71] used genetic data from permafrost to investigate changes in plant and animal diversity during the Pleistocene–Holocene transition, and uncovered evidence pointing to the potential late survival of mammoths and horses until ~9700 years ago in Canada. These results further underscore the dynamic nature of Quaternary ecosystems and demonstrate that palaeoenvironmental DNA can reliably measure biological turnover even without morphologically identifiable specimens.

To date, palaeoenvironmental DNA studies have largely been restricted to either metabarcoding techniques, where hypervariable 'barcoding' regions are amplified using PCR [72,73], or targeting specific markers – frequently mitochondrial genomes – using hybridisation enrichment [68,71]. Recently, some studies have employed **shotgun sequencing** techniques to retrieve genomic data from palaeoenvironmental samples, resulting in true environmental palaeogenomic data, although these have so far mostly been limited to the generation of ancient microbial genomes. For example, Weyrich *et al.* [74] generated partial microbial genomes from >40 000 year old Neandertal **dental calculus**. Comparison of the Neandertal strain of *Methanobrevibacter oralis* – a microbe associated with periodontal disease [75] – to that of modern humans revealed that the strains only diverged 112 000–143 000 years ago, which is significantly more recent than the estimated genomic divergence between these two hominin lineages (450 000–750 000 years ago). This temporal discrepancy provides independent support for hybridisation between Neandertals and modern humans [76], and opens up the possibility of using the oral microbiota to study the evolution of hominins, and perhaps even of other animal taxa. Beyond purely phylogenetic implications, Weyrich *et al.* [74] also showed that genetic differences between the two strains were associated with differences in microbial oral biota, diet, and hygiene in Neandertals and modern humans, suggesting that environmental palaeogenomic data also have the potential to reveal functional information about past ecological communities.

Currently, a major limitation of environmental palaeogenomic studies is that they rely on comparison to sparsely populated nucleotide databases for taxonomic assignment, which in practice means that many sequenced molecules cannot be accurately identified. However, this situation will inevitably improve through time as more modern and palaeogenomic reference data are published. More accurate methods for taxonomic assignment of ancient molecules are also being developed (e.g., [77]), which should maximise the retrieval of usable environmental palaeogenomic data. Another caveat is that current environmental palaeogenomic methods cannot reliably retrieve data of utility in population or phylogenetic analyses that rely on the assignment of sequences to

specific individuals because many individual organisms are likely to contribute to the DNA extracted from an environmental sample. Although in some cases it may be possible to estimate the number of contributing individuals [78], most studies are currently limited to detecting the presence of particular taxa. Nevertheless, detection of a species in the absence of physical remains will be extremely valuable for reconstructing the past distribution and extinction chronology of otherwise rare Quaternary species and is also likely to provide information about taxon-specific preservation potential and taphonomic biases. When combined with palaeogenomic data from single-source bones from the fossil record, environmental palaeogenomic data have the potential to provide a much fuller and more nuanced picture of Quaternary ecosystem change.

Disease Transmission

MacPhee and Marx [79] hypothesised that the Quaternary megafaunal extinctions may have been caused by a disease capable of widespread **zoonotic** transfer across taxonomically diverse animal groups. Indeed, the Quaternary period saw many waves of animal migration – including hominins – often over intercontinental distances, which would have resulted in novel ecological interactions and opportunities for transmission of pathogens and parasites. In addition, during recent history there is some evidence that zoonotic transfer of parasites led to the extinction of island endemics [80]. Nevertheless, although there is little evidence that disease played a role in mass extinctions [81], and the data in fact tend to suggest that direct human impacts and climate change were more influential overall in driving Quaternary extinctions, the role of disease in shaping past ecological communities on a more local scale has not been widely tested. Most palaeogenomic studies of ancient pathogens have focused on the evolution of major diseases in human history, especially bubonic plague [25], and also leprosy [82], gastritis [83], hepatitis B [84], salmonella [85], smallpox [86], and syphilis [87]. By contrast, palaeogenomic studies focusing on past pathogen transfer between animals (including humans) have been limited, although Bos *et al.* [88] demonstrated that *Mycobacterium tuberculosis* genomes from 1000 year old Peruvian human skeletons shared an affinity with those in seals and sea lions, suggesting a transmission mechanism related to past subsistence hunting. In some situations, species extinctions appear to have resulted in the disappearance of endemic pathogens and parasites [89,90], but the role – if any – of disease itself in this process remains elusive. In any case, the dynamic landscape of shifting biotic distributions preserved in the fossil record presents a unique and largely untapped opportunity to test evolutionary hypotheses about new and emerging disease ecosystems through time and may illuminate past extinction processes.

Concluding Remarks and Future Directions

Although true palaeogenomic studies have only been published for a small handful of taxa, where northern hemisphere species are over-represented, they highlight the power that these data hold for augmenting our reconstructions of Quaternary ecological communities. So far, palaeogenomic data have provided insights into natural levels of morphological variation over millennial timescales and how they relate to taxonomic assignment, and have revealed the pervasiveness and impacts of past hybridisation and disease, and recorded demographic trajectories and changes in genomic diversity preceding extinction. Some palaeogenomic studies have even attempted to gain insight into animal behaviour and biology by looking for functional genetic loci that were under natural selection (e.g., [93]).

Although palaeogenomic data offer unique insights into important biological processes and past ecological communities, much work remains to be done for the full benefits of palaeogenomic research to be realised, and our review highlights several key questions that future studies may address (see Outstanding Questions). However, a more dynamic and nuanced picture of past ecosystems is already beginning to emerge. This has implications not only for present-day

Outstanding Questions

How can we more accurately measure past biodiversity changes based on skeletal remains? Palaeogenomic data – although currently only available for a handful of taxa – allow highly accurate inference of the relationships among ancient specimens, allowing quantification of intraspecies morphological variation over multi-millennial timescales. This may provide a guide to interpreting morphological changes in the fossil record, which may alternatively reflect taxonomic turnover or adaptation.

How universal are the biological patterns so far revealed by palaeogenomics? Much of the value of palaeogenomics lies in how the findings can be extrapolated either to the present or the much deeper past, for which we must be confident that the underlying drivers and processes are widely relevant. However, current palaeogenomic studies are heavily biased towards mammalian taxa, particularly those from the Northern Hemisphere (see Figure 1 in main text). Future studies from Australia, Africa, and South America – including birds, reptiles, fish, and plants – will help to fill an important gap.

What are the genetic mechanisms underlying speciation and extinction? In many natural systems, these processes likely play out over very long time-periods, beyond the scope of experimental studies. Palaeogenomic data may allow direct measurement of genes or structural changes that control the evolution of reproductive isolation, as well as the speed at which they emerge, and cast light on the impact of hybridisation and/or long-term gene flow on the speciation process. Similarly, palaeogenomic data may also provide a window into population declines, potentially revealing the consequences of past climate and environmental change.

conservation efforts in the face of anthropogenic biodiversity loss but also – assuming biological processes during the Quaternary are analogous – for inferring the structure and function of ecological communities in the distant past, well beyond the scope of palaeogenomic methods. Important future directions for palaeogenomics will be an expansion to a wider range of study taxa, especially non-domesticated plant species, as well as more intensive use of environmental palaeogenomic records (including coprolites) where identifiable remains of individual organisms are scarce. In addition, the nascent field of **palaeoproteomics** may allow some of the same insights provided by palaeogenomic data to be gleaned from even older time-periods [91,92]. Integration of these data with more traditional palaeontological methods will be crucial for improving the accuracy of our inferences based on the fossil record, which ultimately represents our only source of data for >99% of life's history on Earth.

Acknowledgements

K.J.M. is funded by CABAHC ARC (CE170100015) and the School of Biological Sciences, University of Adelaide; N.J.R. is funded by the Royal Society of New Zealand Marsden Fund (16-UOO-096) and the University of Otago.

References

- Mayr, E. (1997) *This Is Biology: The Science of the Living World*, Harvard University Press
- Higuchi, R. *et al.* (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312, 282–284
- Cooper, A. and Poinar, H.N. (2000) Ancient DNA: do it right or not at all. *Science* 289, 1139
- Knapp, M. *et al.* (2012) Setting the stage – building and working in an ancient DNA laboratory. *Ann. Anat.* 194, 3–6
- Cooper, A. *et al.* (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* 409, 704–707
- Green, R.E. *et al.* (2008) A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134, 416–426
- Noonan, J.P. *et al.* (2005) Genomic sequencing of Pleistocene cave bears. *Science* 309, 597–599
- Green, R.E. *et al.* (2006) Analysis of one million base pairs of Neandertal DNA. *Nature* 444, 330–336
- Poinar, H.N. *et al.* (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 311, 392–394
- Allentoft, M. *et al.* (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* 46, 195–200
- Ramírez, O. *et al.* (2009) Paleogenomics in a temperate environment: shotgun sequencing from an extinct Mediterranean caprine. *PLoS ONE* 4, e5670
- Rohland, N. *et al.* (2010) Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLoS Biol.* 8, e1000564
- Llamos, B. *et al.* (2017) Human evolution: a tale from ancient genomes. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 372, 20150484
- Green, R.E. *et al.* (2010) A draft sequence of the Neandertal genome. *Science* 328, 710–722
- Orlando, L. *et al.* (2013) Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499, 74–78
- Barlow, A. *et al.* (2020) Middle Pleistocene cave bear genome calibrates the evolutionary history of Palaeartic bears. *Curr. Biol.* Published online January 24, 2020. <https://doi.org/10.2139/ssrn.3523359>
- Meyer, M. *et al.* (2017) Palaeogenomes of Eurasian straight-tusked elephants challenge the current view of elephant evolution. *eLife* 6, e25413
- Frantz, L.A.F. *et al.* (2020) Animal domestication in the era of ancient genomics. *Nat. Rev. Genet.* 21, 449–460
- Skoglund, P. *et al.* (2015) Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr. Biol.* 25, 1515–1519
- Heintzman, P.D. *et al.* (2015) Genomic data from extinct North American *Camelops* revise camel evolutionary history. *Mol. Biol. Evol.* 32, 2433–2440
- de Manuel, M. *et al.* (2020) The evolutionary history of extinct and living lions. *Proc. Natl. Acad. Sci. U. S. A.* 117, 10927–10934
- Sackton, T.B. *et al.* (2019) Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* 364, 74–78
- Feigin, C.Y. *et al.* (2018) Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. *Nat. Ecol. Evol.* 2, 182–192
- Feigin, C.Y. *et al.* (2019) Widespread cis-regulatory convergence between the extinct Tasmanian tiger and gray wolf. *Genome Res.* 29, 1648–1658
- Spyrou, M.A. *et al.* (2019) Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. *Nat. Commun.* 10, 4470
- Woodward, H.N. *et al.* (2020) Growing up *Tyrannosaurus rex*: osteohistology refutes the pygmy *Nanotyrannus* and supports ontogenetic niche partitioning in juvenile *Tyrannosaurus*. *Sci. Adv.* 6, eaax6250
- Bunce, M. *et al.* (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* 425, 172–175
- Huynen, L. *et al.* (2003) Nuclear DNA sequences detect species limits in ancient moa. *Nature* 425, 175–178
- Skoglund, P. *et al.* (2013) Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* 40, 4477–4482
- Mittnik, A. *et al.* (2016) A molecular approach to the sexing of the triple burial at the Upper Paleolithic site of Dolní Věstonice. *PLoS ONE* 11, e0163019
- Gower, G. *et al.* (2019) Widespread male sex bias in mammal fossil and museum collections. *Proc. Natl. Acad. Sci. U. S. A.* 116, 19019–19024
- Pečnerová, P. *et al.* (2017) Genome-based sexing provides clues about behavior and social structure in the woolly mammoth. *Curr. Biol.* 27, 3505–3510
- Boyer, P. *et al.* (2018) Molecular resolution to a morphological controversy: the case of North American fossil muskoxen *Bootherium* and *Symbos*. *Mol. Phylogenet. Evol.* 129, 70–76
- Froese, D. *et al.* (2017) Fossil and genomic evidence constrains the timing of bison arrival in North America. *Proc. Natl. Acad. Sci. U. S. A.* 114, 3457–3462
- Heintzman, P.D. *et al.* (2016) Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. *Proc. Natl. Acad. Sci. U. S. A.* 113, 8057–8063
- Shapiro, B. *et al.* (2004) Rise and fall of the Beringian steppe bison. *Science* 306, 1561–1565
- Chang, D. *et al.* (2017) The evolutionary and phylogeographic history of woolly mammoths: A comprehensive mitogenomic analysis. *Sci. Rep.* 7, 44585
- Enk, J. *et al.* (2016) *Mammuthus* population dynamics in Late Pleistocene North America: divergence, phylogeography, and introgression. *Front. Ecol. Evol.* 4, 42

39. Heintzman, P.D. *et al.* (2017) A new genus of horse from Pleistocene North America. *eLife* 6, e29944
40. Perri, A.R. *et al.* (2020) Dire wolves were the last of an ancient New World canid lineage. *Nature* in press
41. Barlow, A. *et al.* (2018) Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.* 2, 1563–1570
42. Westbury, M.V. *et al.* (2020) Hyena paleogenomes reveal a complex evolutionary history of cross-continental gene flow between spotted and cave hyena. *Sci. Adv.* 6, eaay0456
43. Taylor, S.A. and Larson, E.L. (2019) Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* 3, 170–177
44. Petr, M. *et al.* (2020) The evolutionary history of Neanderthal and Denisovan Y chromosomes. *Science* 369, 1653–1656
45. Soubrier, J. *et al.* (2016) Early cave art and ancient DNA record the origin of European bison. *Nat. Commun.* 7, 13158
46. Cahill, J.A. *et al.* (2015) Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol. Ecol.* 24, 1205–1217
47. Cahill, J.A. *et al.* (2018) Genomic evidence of widespread admixture from polar bears into brown bears during the Last Ice Age. *Mol. Biol. Evol.* 35, 1120–1129
48. Miller, W. *et al.* (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2382–E2390
49. Lister, A.M. and Sher, A.V. (2015) Evolution and dispersal of mammoths across the Northern Hemisphere. *Science* 350, 805–809
50. Palkopoulou, E. *et al.* (2018) A comprehensive genomic history of extinct and living elephants. *Proc. Natl. Acad. Sci. U. S. A.* 115, E2566–E2574
51. Slon, V. *et al.* (2018) The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature* 561, 113–116
52. Huerta-Sánchez, E. *et al.* (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512, 194–197
53. Williams, M. and Teixeira, J. (2020) A genetic perspective on human origins. *Biochemist* 42, 6–10
54. Barnes, I. *et al.* (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* 295, 2267–2270
55. Loog, L. *et al.* (2020) Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. *Mol. Ecol.* 29, 1596–1610
56. Sotnikova, M. and Rook, L. (2010) Dispersal of the Canini (Mammalia, Canidae: Caninae) across Eurasia during the Late Miocene to Early Pleistocene. *Quat. Int.* 212, 86–97
57. Sheng, G.-L. *et al.* (2019) Paleogenome reveals genetic contribution of extinct giant panda to extant populations. *Curr. Biol.* 29, 1695–1700
58. Fenderson, L.E. *et al.* (2020) Spatiotemporal landscape genetics: investigating ecology and evolution through space and time. *Mol. Ecol.* 29, 218–246
59. Dietl, G.P. and Flessa, K.W. (2011) Conservation paleobiology: putting the dead to work. *Trends Ecol. Evol.* 26, 30–37
60. Cooper, A. *et al.* (2015) Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. *Science* 349, 602–606
61. Murray, G.G.R. *et al.* (2017) Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science* 358, 951–954
62. Lord, E. *et al.* (2020) Pre-extinction demographic stability and genomic signatures of adaptation in the woolly rhinoceros. *Curr. Biol.* 30, 3871–3879
63. van der Valk, T. *et al.* (2019) Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Curr. Biol.* 29, 165–170
64. Palkopoulou, E. *et al.* (2015) Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* 25, 1395–1400
65. Rogers, R.L. and Slatkin, M. (2017) Excess of genomic defects in a woolly mammoth on Wrangel Island. *PLoS Genet.* 13, e1006601
66. Dussex, N. *et al.* (2019) 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
67. Frankham, R. (2015) Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Mol. Ecol.* 24, 2610–2618
68. Slon, V. *et al.* (2017) Neanderthal and Denisovan DNA from Pleistocene sediments. *Science* 356, 605–608
69. Seersholm, F.V. *et al.* (2020) Rapid range shifts and megafaunal extinctions associated with late Pleistocene climate change. *Nat. Commun.* 11, 2770
70. Murray, D.C. *et al.* (2013) Scrapheap challenge: a novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Sci. Rep.* 3, 3371
71. Murchie, T.J. *et al.* (2020) Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. *Quat. Res.* Published online September 1, 2020. <https://doi.org/10.1017/qua.2020.59>
72. Bálint, M. *et al.* (2018) Environmental DNA time series in ecology. *Trends Ecol. Evol.* 33, 945–957
73. Rawlence, N.J. *et al.* (2014) Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. *J. Quat. Sci.* 29, 610–626
74. Weyrich, L.S. *et al.* (2017) Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* 544, 357–361
75. Horz, H-P. and Conrads, G. (2011) Methanogenic Archaea and oral infections – ways to unravel the black box. *J. Oral Microbiol.* 3, 5940
76. Stringer, C. (2016) The origin and evolution of *Homo sapiens*. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 371, 20150237
77. Cribdon, B. *et al.* (2020) PIA: More accurate taxonomic assignment of metagenomic data demonstrated on sedaDNA from the North Sea. *Front. Ecol. Evol.* 8, 84
78. Seersholm, F.V. *et al.* (2018) Subsistence practices, past biodiversity, and anthropogenic impacts revealed by New Zealand-wide ancient DNA survey. *Proc. Natl. Acad. Sci. U. S. A.* 115, 7771–7776
79. MacPhee, R.D.E. and Marx, P.A. (1997) The 40,000-year plaque: humans, hyperdisease and first contact extinctions. In *Natural Change and Human Impact in Madagascar* (Goodman, S. and Patterson, B., eds), pp. 169–217, Smithsonian Institution Press
80. Wyatt, K.B. *et al.* (2008) Historical mammal extinction on Christmas Island (Indian Ocean) correlates with introduced infectious disease. *PLoS ONE* 3, e3602
81. Kathleen Lyons, S. *et al.* (2004) Was a ‘hyperdisease’ responsible for the late Pleistocene megafaunal extinction? *Ecol. Lett.* 7, 859–868
82. Singh, P. *et al.* (2015) Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. *Proc. Natl. Acad. Sci. U. S. A.* 112, 4459–4464
83. Maixner, F. *et al.* (2016) The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science* 351, 162–165
84. Krause-Kyora, B. *et al.* (2018) Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *eLife* 7, e36666
85. Vågene, Å.J. *et al.* (2018) *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* 2, 520–528
86. Mühlemann, B. *et al.* (2020) Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. *Science* 369, eaaw8977
87. Schuenemann, V.J. *et al.* (2018) Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe. *PLoS Pathog.* 14, e1006997
88. Bos, K.I. *et al.* (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514, 494–497
89. Boast, A.P. *et al.* (2018) Coprolites reveal ecological interactions lost with the extinction of New Zealand birds. *Proc. Natl. Acad. Sci. U. S. A.* 115, 1546–1551
90. Lafferty, K.D. and Hopkins, S.R. (2018) Unique parasite aDNA in moa coprolites from New Zealand suggests mass parasite extinctions followed human-induced megafauna extinctions. *Proc. Natl. Acad. Sci. U. S. A.* 115, 1411–1413
91. Presslee, S. *et al.* (2019) Palaeoproteomics resolves sloth relationships. *Nat. Ecol. Evol.* 3, 1121–1130
92. Chen, F. *et al.* (2019) A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. *Nature* 569, 409–412

93. Barnett, R. *et al.* (2020) Genomic adaptations and evolutionary history of the extinct scimitar-toothed cat, *Homotherium latidens*. *Curr. Biol.* Published online October 15, 2020. <https://doi.org/10.1016/j.cub.2020.09.051>
94. Marciniak, S. *et al.* (2020) Evolutionary and phylogenetic insights from a nuclear genome sequence of the extinct, giant 'subfossil' koala lemur *Megaladapis edwardsi*. *bioRxiv* Published online October 17, 2020. <https://doi.org/10.1101/2020.10.16.342907>
95. Chen, N. *et al.* (2020) Ancient genomes reveal tropical bovid species in the Tibetan Plateau contributed to the prevalence of hunting game until the late Neolithic. *Proc. Natl. Acad. Sci.* 117, 28150–28159
96. Ramos-Madrigal, J. *et al.* (2020) Genomes of Pleistocene Siberian wolves uncover multiple extinct wolf lineages. *Curr. Biol.* Published online October 29, 2020. <https://doi.org/10.1016/j.cub.2020.10.002>