

Ancient DNA as a tool for medical research

Gaspard Kerner, Jeremy Choin & Lluís Quintana-Murci

 Check for updates

Paleogenomics can help elucidate the genetic basis of modern diseases, including inborn errors of immunity that impair the response to infections, providing a tool for drug development.

The advent of new technologies and analytical approaches in genomics, such as high-throughput sequencing and CRISPR–Cas gene editing, has opened new avenues of medical research. Polygenic risk scores are being assessed for routine use by the UK National Health Service, and the use of gene editing should improve outcomes for many patients. Similarly, improvements in techniques for obtaining and sequencing DNA from fossil remains have been a major breakthrough in the field of evolutionary genetics. These discoveries have shed new light on the origins of human populations, their migratory history, and the extent of admixture between humans and ancient, now-extinct hominins, such as Neanderthals, and between modern human populations^{1–3}. However, ancient genomics is also of value for medical research, making it possible to reconstruct the history of human health over time, including past epidemics. This research allows increased understanding of the present-day links between genomic diversity and disease.

A golden age of paleogenomics

The study of ancient DNA (aDNA) has developed into a field in its own right: paleogenomics. The oldest nuclear genome sequences from the genus *Homo* obtained thus far were from a 430,000-year-old Neanderthal⁴, but most available ancient human genomes date back no more than 10,000 years (Table 1). aDNA has been extracted from samples from around the world, but particularly from the high latitudes of the northern hemisphere, reflecting an archeological bias intrinsic to the

preservation of samples but also a more general bias toward the investigation of people of European ancestry in genomic studies⁵.

Paleogenomics can contribute to studies of human physiology in obvious ways, through the identification of regions of the human genome affected by the introgression of Neanderthal material, for example, which has revealed genes of major physiological relevance^{2,3}. However, the potential contributions of paleogenomics to medical research are less self-evident.

The increasing availability of aDNA samples is making it possible to address key questions related to human health and disease. For example, humans are an extremely successful species, but little is known about how they survived life-threatening environmental pressures, such as exposure to pathogens, in the past. Research on modern-day humans has shown that specific DNA mutations that alter host defense mechanisms can largely explain differential outcomes after infection. Thus, studying fluctuations in the frequency of genetic variants associated with infectious-disease risks in fossil remains, which represent a continuous and uncontrolled natural experiment, provides an obvious example of the potential value of paleogenomics in medicine.

Immune disorders

The simplest model used to study human susceptibility to infection is genetic predisposition to infectious diseases through inborn errors of immunity – mutations that increase the risk of severe infections⁶. The most successful approach has been to study people who are extremely susceptible to infections, as they have the highest odds of carrying highly penetrant genetic lesions. The main advantage of this approach is that causality between genotype and phenotype follows naturally from the study of an *in vivo* (human) model, although its implementation requires the extensive genetic screening of severely ill patients.

An alternative approach is to study the effects of natural selection from pathogenic pressure on human genome variability⁷. These two



Table 1 | Spatiotemporal distribution of publicly available aDNA data

	>45,000	45,000–10,000	10,000–5,000	5,000–3,000	3,000–1,000	1,000–500	≤500
Africa	0	10	18	16	149	11	12
Europe	2	72	1,099	1,829	2,072	515	48
Asia	0	62	452	1,048	1,173	185	63
Oceania	0	0	0	0	56	52	56
North America	0	7	18	41	122	226	76
South America	0	5	33	7	52	69	39

The number of humans with whom retrieved aDNA sequences are associated is shown. Data were retrieved from the Allen Ancient DNA (AADR), data release V54.1, accessed 15 December 2022 (<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>). Time intervals (columns) are given in years before 1950 AD, as is common practice in paleogenomics.

frameworks are comparable in that both involve the identification of variants that increase the risk of a given infectious disease in natura. However, they differ in that inborn errors of immunity usually operate at the scale of a single generation, whereas gradual pathogenic pressure operates at the scale of many generations, which makes it possible to identify genetic variants with effect sizes that differ by several orders of magnitude.

Recent studies have highlighted the value of using ancient genomes from different epochs, known as aDNA time series, to reconstruct the evolutionary history of immune disorders and past epidemics (Fig. 1). One recent proof-of-concept study of more than 1,000 genomes dated to within the last 10,000 years of European history showed that a tuberculosis risk variant, *TYK2* P1104A, present in around 3% of people of European ancestry, has evolved under strong negative selection over the past two millennia⁸. This finding, probably reflective of the pressure imposed by *Mycobacterium tuberculosis*, would have been very difficult to achieve through studies of modern DNA. Indeed, most methods for detecting natural selection in modern DNA data are underpowered for low-frequency variants.

The evolutionary history of the pathogen itself can also provide insight into the dynamics of past epidemics, but this was difficult to characterize until samples from ancient pathogens became available. In the context of tuberculosis, studies of ancient mycobacteria have dated the most recent common ancestor of the pathogen to only 6,000 years ago, which contrasts markedly with the estimates of more than 70,000 years ago that were obtained from studies of modern strains of *M. tuberculosis*⁹. Medical practitioners may find aDNA studies similarly helpful for pinpointing genetic variants of microorganisms, as their evolutionary history can reveal their deleteriousness to human health.

aDNA time series at the scale of the entire human genome can identify variants under negative selection, which is useful for the detection of new genetic factors associated with immune disorders, as described in a recent study¹⁰. Paleogenomics therefore appears to be a powerful approach, complementary to epidemiological and clinical genetics studies, that can be used to confirm and expand on genetic variants associated with disease risk.

Gradual adaptation to pathogens

aDNA time series studies can also be used to identify advantageous mutations in humans that are positively selected over time through pathogen exposure. Such approaches have shown that protective variants in immunity-related genes, including the *TLR10–TLR1–TLR6* gene cluster and the human leukocyte antigen-associated *PPT2* and *EGFL8*, have been recurrent targets of positive selection over the past

10,000 years^{11,12}, particularly since the start of the Bronze Age in Europe ~4,500 years ago^{8,10}.

Some signals of positive selection can be delineated in their historical context. For example, some immunity-related genes under positive selection have been found to be associated with the Neolithic transition in Europe. This is the case for *FUT2*, which encodes a molecule involved in determining ABO blood group status; *IL1R2*, for which high levels of expression are associated with protection against several autoimmune disorders; and *CACNB1*, which encodes a T cell regulator¹¹. Collectively, paleogenomic studies using aDNA time transects provide a framework for the identification of genes that have both participated in human adaptation to past epidemics and shaped the immune system of modern human populations.

Legacy of now-extinct humans

Admixture between modern humans and now-extinct hominins, such as Neanderthals or Denisovans, may have provided modern humans with a selective advantage through the exchange of genetic material. Immunity-related genetic variants introgressed from archaic hominins into the genomes of modern humans can inform host resistance to pathogens and serve as candidate drug targets in disease treatment. As most archaic genetic material has undergone purifying selection in modern humans, the archaic variants that have survived in the human genome probably perform key functions for human survival in general and for pathogen resistance in particular. For example, several studies have identified high levels of Neanderthal ancestry for genes encoding molecules involved in innate immunity, genes encoding RNA-virus-interacting proteins and genetic variants associated with variations in gene expression after viral stimulation^{13–15}. A particularly interesting example is provided by a Neanderthal haplotype encompassing the antiviral *OAS1* gene, which was recently shown to be associated with protection against severe COVID-19¹⁶.

There is also growing evidence to suggest that Denisovan variants introgressed into modern human genomes have affected mainly immunity-related functions^{17,18}. However, unlike Neanderthal ancestry, which is homogeneously distributed between non-African populations, the Denisovan heritage is highly variable across modern human populations, being especially prevalent in populations from the Asia-Pacific region. Like studies of aDNA time transects, analyses of archaic introgression can help to elucidate the beneficial roles of certain genes in human immunity.

Pathogen-driven selection

Studies of genetic adaptation to pathogens cannot, in general, be linked to specific pathogens. An iconic example illustrating this caveat is

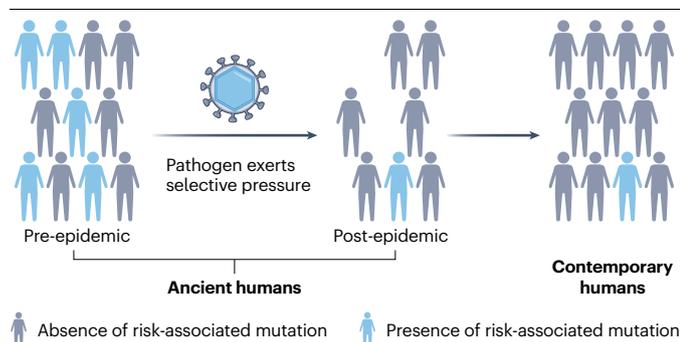


Fig. 1 | Ancient DNA can identify genetic variants associated with disease risk. Schematic representation of the use of aDNA to study the effects of negative selection on genetic variants associated with disease risk in the context of past pathogenic pressure. Here, the aDNA samples date from either before or after an epidemic event. The frequencies of DNA variants between the two groups of samples (pre- and post-epidemic groups) can identify genetic variants that are targeted by negative selection, which are present at a significantly lower frequency than would be expected by chance in the post-epidemic group. Such observations provide clues to the pathogenic nature of the variant, and the corresponding gene, in the context of the infectious disease studied.

the selection pressures detected for *CCR5*¹⁹, for which many polymorphisms, in particular a 32-base-pair deletion in the open reading frame of the gene, have been associated with differential outcomes after infection with human immunodeficiency virus type 1. The selection signatures observed at the *CCR5* locus cannot be attributed to this virus, which emerged much too recently to have already left a selection footprint in our genomes; they must, therefore, result from pressures exerted by other pathogens in the past.

However, recent paleogenomic studies have proved informative for the reconstruction of past epidemics caused by specific infectious agents. The most relevant example is that of plague and its causal agent, *Yersinia pestis*. Three historical plague pandemics are known to have occurred, but it was not until the recovery of aDNA from *Y. pestis* strains that the bacterium originally associated with the last pandemic of the nineteenth to twentieth centuries was also found to be the causal agent of the Justinian plague in the sixth century and the Black Death in the fourteenth century²⁰. The ability of paleogenomics to shed light on human genetic factors associated with disease risk is illustrated by a recent study comparing aDNA samples from humans who lived before and after the Black Death²¹. The authors identified a specific genetic variant (rs2549794) close to the *ERAP2* gene that seems to be associated with greater protection against plague. This variant has a strong signature of positive selection, with a selective coefficient of 0.4, which means that people who were homozygous for this protective allele were 40% more likely to survive and reproduce during the Black Death pandemic than those homozygous for the non-protective allele, and macrophages expressing this allele have a greater capacity to control *Y. pestis* than those expressing the non-protective allele. Understanding present-day susceptibility to disease through aDNA will pave the way for improvements in personalized disease prevention and treatment.

Future priorities

With the increasing availability of high-quality aDNA samples, expectations are high for the future of paleogenomics and its contribution to medical research. However, most samples available thus far were genotyped with an in-solution enrichment technique targeting about

1.2 million frequent variants of the genome. Use of the alternative technique of shotgun sequencing would facilitate the detection of variants under negative selection that have become rare (or even extinct) in current populations. Such studies would allow the identification of novel genes containing variants deleterious for human health and associated with disease risk.

The same gene can underlie different phenotypic traits, a phenomenon known as pleiotropy, so the success of gene-targeted drug development depends on a careful assessment of potential collateral damage. One recent study of more than 2,000 available European ancient genomes showed that variants that alter both the risk of infection and autoimmune traits have been a chief substrate of selection in recent millennia¹⁰. The prevalence of variants that increase the risk of autoimmune disease has grown over time, most likely as the result of positive selection due to pathogenic pressure, as these same variants reduce the risk of infectious disease. The identification of such antagonistic pleiotropic variants acting during evolution should facilitate strategies for drug development that minimize the risk of secondary effects.

Sequencing of the gut microbiota over time can also provide information about infection. For example, Wibowo and colleagues compared DNA extracted from fossilized stools dating back to 1,000–2,000 years ago with that of stool samples from present-day populations with industrialized or non-industrialized lifestyles²². Unsurprisingly, the composition and diversity of ancient gut microbiota were closer to those of non-industrialized present-day populations, but the ancient microbes lacked the antibiotic-resistance-related genes present in both industrialized modern populations and non-industrialized modern populations. This illustrates how the study of ancient microbes could provide insight into the spread and evolution of antibiotic resistance – one of the greatest threats to global health today.

Finally, the sequencing of ancient proteins, such as antibodies, from human remains might better account for past acute infections than sequencing DNA from the pathogens. Ancient protein research should also improve knowledge of the evolution of human immunity, host–pathogen interactions and the dynamics of past epidemics. It is increasingly clear that paleogenomics is a discipline of interest well beyond purely anthropological questions, as it can provide answers to questions of fundamental importance in medical research. It may be time to adopt a slightly rephrased version of Theodosius Dobzhansky's famous phrase: “Nothing in medicine makes sense except in the light of evolution.”

Gaspard Kerner¹, **Jeremy Choin**² & **Lluis Quintana-Murci**^{1,3}✉

¹Human Evolutionary Genetics Unit, Institut Pasteur, Université Paris Cité, Centre National de la Recherche Scientifique (CNRS) UMR, 2000 Paris, France. ²Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. ³Chair of Human Genomics and Evolution, Collège de France, Paris, France.

✉ e-mail: quintana@pasteur.fr

Published online: 15 March 2023

References

- Skoglund, P. & Mathieson, I. *Annu. Rev. Genomics Hum. Genet.* **19**, 381–404 (2018).
- Gittelman, R. M. et al. *Curr. Biol.* **26**, 3375–3382 (2016).
- Dannemann, M. & Racimo, F. *Curr. Opin. Genet. Dev.* **53**, 1–8 (2018).
- Meyer, M. et al. *Nature* **531**, 504–507 (2016).
- Sirugo, G. et al. *Cell* **177**, 26–31 (2019).
- Casanova, J.-L. & Abel, L. *Cell* **185**, 3086–3103 (2022).
- Quintana-Murci, L. *Cell* **177**, 184–199 (2019).

8. Kerner, G. et al. *Am. J. Hum. Genet.* **108**, 517–524 (2021).
9. Bos, K. I. et al. *Nature* **514**, 494–497 (2014).
10. Kerner, G. et al. *Cell Genomics* **3**, 100248 (2023).
11. Le, M. K. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.08.24.505188> (2022).
12. Mathieson, I. et al. *Nature* **528**, 499–503 (2015).
13. Deschamps, M. et al. *Am. J. Hum. Genet.* **98**, 5–21 (2016).
14. Quach, H. et al. *Cell* **167**, 643–656.e617 (2016).
15. Enard, D. & Petrov, D. A. *Cell* **175**, 360–371.e313 (2018).
16. Zeberg, H. & Pääbo, S. *Proc. Natl Acad. Sci. USA* **118**, e2026309118 (2021).
17. Choin, J. et al. *Nature* **592**, 583–589 (2021).
18. Vespasiani, D. M. et al. *PLoS Genet.* **18**, e1010470 (2022).
19. Bamshad, M. J. et al. *Proc. Natl Acad. Sci. USA* **99**, 10539–10544 (2002).
20. Spyrou, M. A. et al. *Nat. Rev. Genet.* **20**, 323–340 (2019).
21. Klunk, J. et al. *Nature* **611**, 312–319 (2022).
22. Wibowo, M. C. et al. *Nature* **594**, 234–239 (2021).

Acknowledgements

The Human Evolutionary Genetics Laboratory is supported by the Institut Pasteur, the Collège de France, the CNRS, Agence Nationale de la Recherche (ANR) grants COVID-19-POPCELL (ANR-21-CO14-0003-01), POPCELL-REG (ANR-22-CE12-0030-01), COVIFERON (ANR-21-RHUS-0008), LIFECHANGE (ANR 17 CE12 0018 02), CNSVIRGEN (ANR-19-CE15-0009-02) and MORTUI (ANR-19-CE35-0005), the EU HORIZON-HLTH-2021-DISEASE-04-07 grant UNDINE (no. 101057100), the French Government's Investissement d'Avenir program, Laboratoires d'Excellence "Integrative Biology of Emerging Infectious Diseases" (ANR-10-LABX-62-IBEID) and "Milieu Intérieur" (ANR-10-LABX-69-01), the Fondation pour la Recherche Médicale (Equipe FRM DEQ20180339214), the Fondation Allianz-Institut de France and the Fondation de France (no. 00106080). G.K. is supported by a Pasteur-Roux-Cantarini fellowship.

Competing interests

The authors declare no competing interests.