CellPress

Cell
Leading Edge

Commentary

Evolving ancient DNA techniques and the future of human history

Yichen Liu,[1,2] E. Andrew Bennett,[1,2] and Qiaomei Fu[1,2,3,*]
[1]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 100044, China
[2]Shanghai Qi Zhi Institute, Shanghai 200232, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
*Correspondence: fuqiaomei@ivpp.ac.cn
https://doi.org/10.1016/j.cell.2022.06.009

Ancient DNA (aDNA) techniques applied to human genomics have significantly advanced in the past decade, enabling large-scale aDNA research, sometimes independent of human remains. This commentary reviews the major milestones of aDNA techniques and explores future directions to expand the scope of aDNA research and insights into present-day human health.

Introduction of NGS techniques in aDNA field

The abundance of paleogenomic data of ancient humans has been accumulating rapidly since 2010. Over six thousand reconstructed ancient human genomes have allowed detailed insights into the migrations, interactions, expansions, and disappearances of past human populations that were previously inaccessible. One milestone that paved the way to such accomplishments was the introduction of next-generation sequencing (NGS) techniques in the ancient DNA (aDNA) field. The application of NGS on ancient human remains resulted in not only a vast increase in the quantity of bases generated compared to the pre-NGS era, but importantly, reduced the impact trace amounts of modern DNA had on the reliability of results, a problem that had plagued the previous amplification-based approaches using PCR. NGS methods also allowed for more detailed taphonomic characterization of aDNA molecules, such as fragmentation from depurination and C-to-T substitutions from cytosine deamination. Studies such as these gave rise to the subsequent gold standard criteria for aDNA authentication and permitted the separation *in silico* of endogenous from contaminating DNA, which overcame a major challenge of the pre-NGS era when researchers relied on the independent replication of results and *a priori* phylogenetic expectations to support the authenticity of ancient human DNA.

The ability to reconstruct ancient human genomes made of the genomic information of past humans has not only allowed us to learn the timing and pathways of past human migrations, but also made the genomes of two extinct archaic human lineages, the Neanderthals and Denisovans, accessible for the first time. Neanderthals and Denisovans appear to be the sister groups of modern humans (*homo sapiens*) that had separated from a common ancestor roughly 550,000 years ago. The first Neanderthal and Denisovan genomes documented interactions between each other and with modern humans for an extended time period. The traces of Neanderthal DNA are still detectable in present-day non-Africans, and Denisovan DNA is still present in Oceanian, East Asian, and American genomes. Additionally, both the Y chromosomes and mitochondrial genomes of Neanderthals may have been replaced by those of a modern human lineage more than 200,000 years ago (Liu et al., 2021).

Over the past decade, NGS techniques have been continuously optimized for better performance specific to aDNA applications (Orlando et al., 2021; Figure 1). In addition to widely recognized higher rates of aDNA survival in petrous bones, a major innovation in aDNA library preparation is uracil DNA glycosylase (UDG) treatment. This process reduces the effect of post-mortem cytosine deamination of the aDNA molecules by removing deaminated cytosines (uracils). To use a single library for both authentication of the endogenous content and accurate calling of the ancient variants, a modified protocol—partial UDG treatment—has been developed to remove DNA damage (i.e., C- > T) in the interior of DNA molecules, while partially keeping the original damaged state at the terminal ends of the molecule (first base on 5′ or 3′ ends) (Rohland et al., 2015). Another major development specific to improving the recovery of aDNA was the single-stranded DNA library construction protocol (Gansauge and Meyer, 2013). aDNA, especially from poorly preserved samples, contains nicks, gaps, and single-stranded overhangs, elements that cannot be efficiently sequenced when using a double-stranded DNA library preparation protocol. Single-stranded library preparation overcomes these inefficiencies by ligating adapters directly to denatured single-stranded DNA molecules, allowing the recovery of DNA molecules that would have been lost using double-stranded protocols. This advance increased library yields from aDNA by approximately one order of magnitude and was used to generate the first high coverage (>30×) paleolithic genomes of ancient humans, opening the doorway for precision in the genomic analysis of past populations on par with that of modern population studies. Also, upstream sample preparation and aDNA extraction and downstream bioinformatics steps have seen considerable improvements: including removing trace contaminants using bleach, optimizing binding buffers to recover increasingly short DNA fragments, as well as algorithms and parameters of

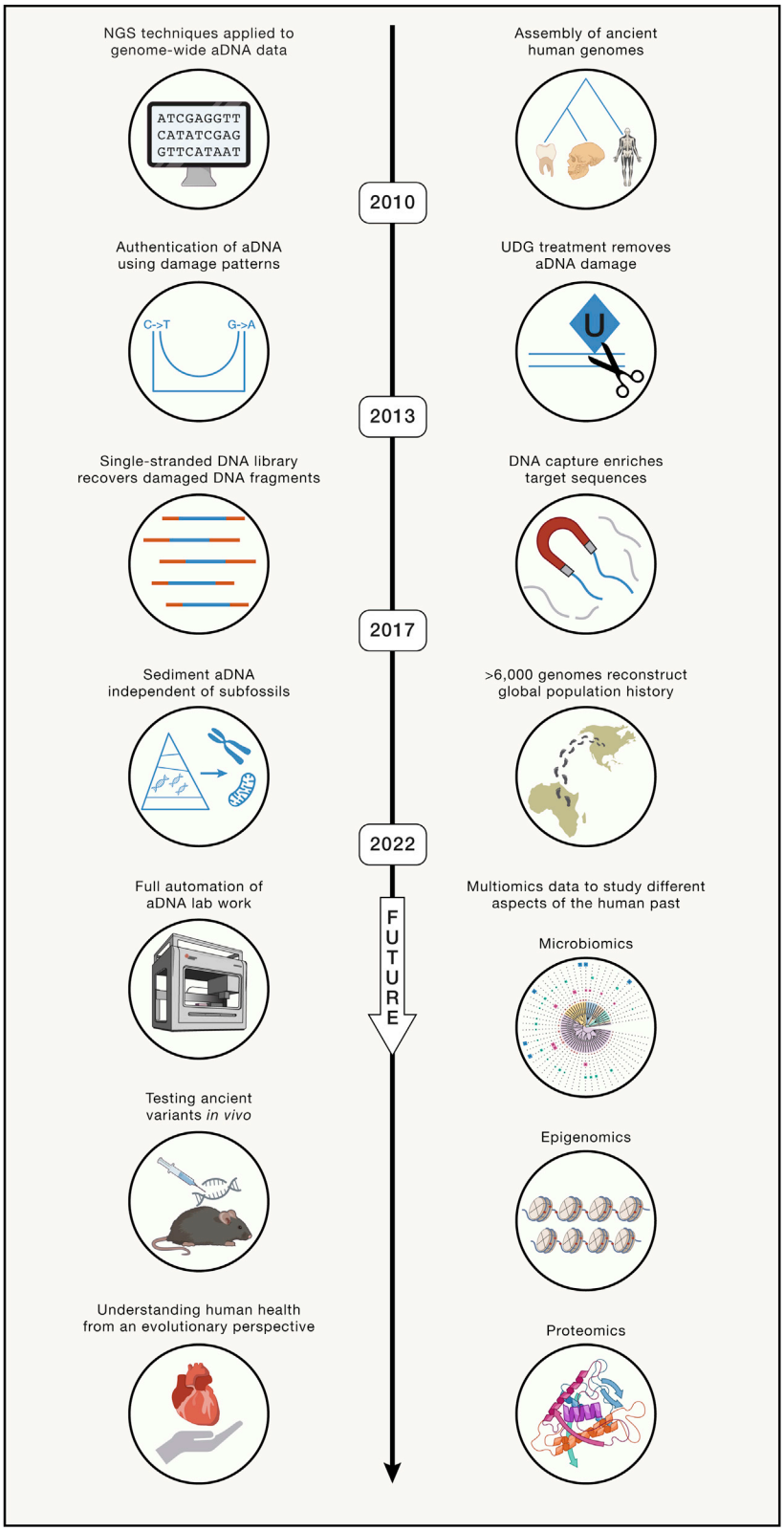

Figure 1. Schematic figure illustrating advancements in aDNA techniques since 2010

read alignment and single nucleotide polymorphism (SNP) calling for aDNA characteristics (Korlević and Meyer, 2019; Prüfer, 2018).

DNA capture techniques increase the accessibility of aDNA information

Frequently, the endogenous DNA in an ancient sample makes up fewer than 1% of the sequenced reads, multiplying shotgun sequencing costs. As a consequence, the application of shotgun sequencing to a large number of samples is impractical for recovering the specific regions needed to make genome-to-genome comparisons of partially sequenced genomes. One solution is to use "DNA capture," which makes use of short DNA or RNA oligo "baits" that hybridize with target sequences to enrich for specific genomic regions of interest from multiple samples. For ancient human studies, several panels of baits overlapping collections of phylogenetically or biologically informative SNPs have been designed to reflect the genetic variations among past and present human populations (Orlando et al., 2021). These panels are widely used in aDNA studies, and about two-thirds of currently available ancient-human-genome-wide data are generated using the same panel, covering roughly 1,240,000 SNPs.

The application of DNA capture techniques has increased the scale of aDNA studies and expanded the time and geographical range covered by ancient human genomes. For example, genomic data of ancient humans from Africa and southern East Asia, areas important for understanding human population history but where heat or humidity are unfavorable for aDNA preservation, have been successfully recovered using DNA capture techniques. Recent ancient African data revealed geographically clustered lineages that reflect the distribution of modern human populations on the African continent, and the ancient genomes of southern East Asians revealed interactions among populations living in southern East Asia and Southeast Asia that were more complex than previously inferred using modern genomic or archeological data (Liu et al., 2021).

However, the use of human subfossils for aDNA study poses a limitation. The record of human physical remains contains

large gaps in time and geography, and the human subfossils that have been discovered are scarce and irreplaceable resources, which constrains the samples available for destructive DNA analysis. A recent breakthrough in the aDNA field extended the application of DNA capture techniques from subfossils to sediments. Slon et al. (2017) were able to retrieve ancient mammalian mtDNA, including human, directly from Pleistocene cave sediments. The utility of sediments as a source for aDNA overcomes a major obstacle of the limited availability of physical subfossil samples and may potentially have a major impact on our ability to reconstruct the past through ancient genetics. In 2020, Denisovan mtDNA was obtained from sediments in Baishiya Karst Cave, which was the first genetic evidence of Denisovans obtained outside Denisova Cave in Siberia, and may help to explain the origin of the Denisovan genetic adaptation to high altitudes that later appears in some East Asians (Zhang et al., 2020). Compared to the high copy number and small genome size of mtDNA, the more challenging human nuclear DNA was only recently obtained from sediments. A probe set targeting 1.6 million archaic-human-specific SNPs has been applied to study cave deposits and revealed a Neanderthal population turnover that occurred ~100,000 years ago in northern Spain (Vernot et al., 2021). Even though difficulties in identifying the number of individuals contributing genomic DNA to a given sediment sample make this technique more useful for studies of past populations than individuals, the beginning of a genomic era of sediment aDNA research is well underway and shows great promise.

Future perspectives

Full automation of aDNA lab work

Despite unprecedented achievements in our ability to mine high-quality ancient genomic information from multiple samples with varied preservation status, current paleogenomic techniques are still far from sufficient to offer a thorough understanding of human genetic history. Detailed characterization of the complexity of past populations cannot be achieved with small collections of individuals spread over thousands of years and large geographical areas. An important future step in aDNA techniques will be to rapidly process large numbers of aDNA samples. The inherent characteristics of aDNA require screening large numbers of samples and labor-intensive experiments before obtaining sufficient data for bioinformatic analyses. This is especially true in sediment studies, with a high variability of aDNA recovery from different layers and living areas. For example, in a cave with direct aDNA evidence confirming the extended presence of archaic humans, hominin DNA was found in only 15 out of 87 sediment samples, and 102 DNA libraries were subsequently constructed from 10 samples to obtain reads sufficient for analysis (Slon et al., 2017).

Currently, several aDNA laboratories have turned to liquid-handling robots, which dramatically reduce the hands-on time needed for DNA extraction, library preparation, and capture protocols, as well as reducing the potential for contamination and human error when manually processing large sample numbers. Still, one step remains to be automated, that of sample preparation, which is a labor-intensive process including the cleaning and pulverization of ancient samples (e.g., teeth and bones) and the transfer of these to extraction buffers. Automation-assisted large-scale aDNA profiling is appealing for multiple reasons. One application is to process a large number of sediments from archaeological sites. Additionally, some large historical or prehistorical sites, such as large-scale cemeteries or the remains of ancient battlefields, may preserve hundreds or even thousands of human remains. Genomic profiling with kinship analysis of such ancient societies can reveal insights into demographic structure and its association with past sociocultural organization. In fact, prehistoric kinship practice and social inequality have already been revealed by kinship-based analysis on dozens of individuals, exhibiting the great potential of detailed genetic profiling of past societies (Fowler et al., 2022). All of these studies would benefit from increased scalability from the implementation of automated procedures.

Extend the scope beyond ancient human DNA

While the sequences of genomes of ancient humans are informative regarding their genetic, demographic, and evolutionary history, a more comprehensive understanding can be gained from other types of aDNA information, such as microbial and epigenetic information (Orlando et al., 2021; Figure 1).

Much of our understanding of the lifestyle, diets, and morphology of archaic humans is currently limited to genomic and archaeological evidence. The study of aDNA from ancillary sources can gain additional insights into these areas. For example, the metagenomic DNA obtained from ancient dental calculus can capture the dietary and oral microbial DNA of ancient humans. Microbial pathogen DNA obtained from ancient bones possesses the potential to reveal past epidemics and pandemics, although DNA capture is usually required to obtain pathogen DNA due to its lower proportion compared to host DNA. Further, while the morphology of Denisovans still remains largely unknown due to the scarcity of identified skeletal remains, epigenetic information has been leveraged for inferring phenotypic traits of Denisovans; an elongated face and a wide pelvis have been predicted from their methylomes.

Other molecular sources, such as RNA and proteins, have also been retrieved successfully from ancient samples (Orlando et al., 2021). Ancient RNA has the potential to investigate past human transcriptomes, although the instability of RNA over time has made successful ancient RNA studies extremely limited and ancient human RNA has yet to be retrieved. On the other hand, ancient protein can be preserved longer than DNA and thus is a good candidate for studying extremely ancient samples, allowing us to travel further into the past. Also, Zooarchaeology by Mass Spectrometry (ZooMS) is a cost-effective proteomic method primarily used for taxonomic identification. Thus, taxonomically characterizing ancient proteins using ZooMS can serve as a screening strategy to identify non-diagnostic bone fragments. Brown et al. (2022) analyzed ~3,800 morphologically ambiguous bone fragments from Denisova Cave using peptide mass fingerprinting and identified five hominin samples, from which the oldest Denisovan DNA to date (~200 kya) has been recovered.

Understanding present-day human health from an evolutionary perspective

Genomic regions that have undergone selection have been identified with help from

the increasing knowledge of archaic and ancient human genomes. Several functionally important haplotypes that involve innate immunity, lipid metabolism, adaptation to high altitude, and skin pigmentation (Racimo et al., 2015), as well as the recently revealed haplotypes that determine susceptibility to severe COVID-19, have been suggested to be inherited from archaic humans (Racimo et al., 2015; Zeberg and Pääbo, 2021). Within modern human populations from different time periods, for example before and after the Last Glacial Maximum, changes in frequencies of haplotypes related to phenotypes like skin pigmentation, eye color, and tooth morphology and hair thickness (Fu et al., 2016) are observed, some of which may have been adaptations to climate changes. A further step would be to precisely identify potential regions of evolutionary importance and verify their functional consequences *in vivo* using gene-editing techniques like CRISPR-Cas9, as has been done with the archaic variant of *NOVA1* that was revealed to be associated with the development of cortical organoids (Trujillo et al., 2021). In the future, high-quality ancient human genomes over an evolutionarily important time scale and the establishment of experimental systems including cell lines, organoids, and animal models, will facilitate verification of the phenotypic consequences of genomic variants. These systems could be further expanded to verify epigenomic and microbiomic signals (Figure 1).

Focus on improving decontamination, DNA extraction, and bioinformatics techniques tailored to the characteristics of ancient molecules continues to extend reliable analysis to more poorly preserved samples. The success in recovering the DNA of ancient humans from sediment has also unlinked, to some degree, the dependence on rare archaeological specimens for DNA extraction. Furthermore, methodologies developed in the aDNA field have increasingly been applied to applications beyond the study of the past. The ability to accurately retrieve trace amounts of DNA and the understanding of contamination dynamics can be valuable in the field of forensics, research involving low biomass samples like airway microbes, single-cell genomics, and the study of cell-free DNA. Moving forward, with the continued development and improvement of aDNA techniques, we can expect an increase in the rate of insights into the evolutionary and demographic history of humans, as well as a broader application of these techniques in the near future.

ACKNOWLEDGMENTS

We acknowledge constructive discussions and feedback from Molecular Paleontology Laboratory members at IVPP. This work was supported by the Chinese Academy of Sciences (CAS) and the Ministry of Finance of the People's Republic of China (YSBR-019 and XDB26000000), the National Natural Science Foundation of China (41925009 and T2192950), the Tencent Foundation (through the XPLORER PRIZE), the Howard Hughes Medical Institute (grant 55008731), and the Shanghai Qi Zhi Institute.

AUTHOR CONTRIBUTIONS

Conceptualization: Q.F.; writing, original draft: Y.L., E.A.B., and Q.F.; reviewing and editing: Y.L., E.A.B., and Q.F.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Brown, S., Massilani, D., Kozlikin, M.B., Shunkov, M.V., Derevianko, A.P., Stoessel, A., Jope-Street, B., Meyer, M., Kelso, J., Pääbo, S., et al. (2022). The earliest Denisovans and their cultural adaptation. Nature ecology & evolution *6*, 28–35. https://doi.org/10.1038/s41559-021-01581-2.

Fowler, C., Olalde, I., Cummings, V., Armit, I., Büster, L., Cuthbert, S., Rohland, N., Cheronet, O., Pinhasi, R., and Reich, D. (2022). A high-resolution picture of kinship practices in an Early Neolithic tomb. Nature *601*, 584–587. https://doi.org/10.1038/s41586-021-04241-4.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of ice age Europe. Nature *534*, 200–205. https://doi.org/10.1038/nature17993.

Gansauge, M.-T., and Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat. Protoc. *8*, 737–748. https://doi.org/10.1038/nprot.2013.038.

Korlević, P., and Meyer, M. (2019). Pretreatment: removing DNA contamination from ancient bones and teeth using sodium hypochlorite and phosphate. In Ancient DNA (Springer), pp. 15–19.

Liu, Y., Mao, X., Krause, J., and Fu, Q. (2021). Insights into human history from the first decade of ancient human genomics. Science *373*, 1479–1484. https://doi.org/10.1126/science.abi8202.

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P.W., Ávila-Arcos, M.C., Fu, Q., Krause, J., Willerslev, E., Stone, A.C., and Warinner, C. (2021). Ancient DNA analysis. Nature Reviews Methods Primers *1*, 14–26. https://doi.org/10.1038/s43586-020-00011-0.

Prüfer, K. (2018). snpAD: an ancient DNA genotype caller. Bioinformatics *34*, 4165–4171. https://doi.org/10.1093/bioinformatics/bty507.

Racimo, F., Sankararaman, S., Nielsen, R., and Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. Nat. Rev. Genet. *16*, 359–371. https://doi.org/10.1038/nrg3936.

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Phil. Trans. Biol. Sci. *370*, 20130624. https://doi.org/10.1098/rstb.2013.0624.

Slon, V., Hopfe, C., Weiß, C.L., Mafessoni, F., De La Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M.V., Miller, R., et al. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. Science *356*, 605–608. https://doi.org/10.1126/science.aam9695.

Trujillo, C.A., Rice, E.S., Schaefer, N.K., Chaim, I.A., Wheeler, E.C., Madrigal, A.A., Buchanan, J., Preissl, S., Wang, A., Negraes, P.D., et al. (2021). Reintroduction of the archaic variant of NOVA1 in cortical organoids alters neurodevelopment. Science *371*, eaax2537. https://doi.org/10.1126/science.aax2537.

Vernot, B., Zavala, E.I., Gómez-Olivencia, A., Jacobs, Z., Slon, V., Mafessoni, F., Romagné, F., Pearson, A., Petr, M., Sala, N., et al. (2021). Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. Science *372*, eabf1667. https://doi.org/10.1126/science.abf1667.

Zeberg, H., and Pääbo, S. (2021). A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. Proc. Natl. Acad. Sci. USA *118*. https://doi.org/10.1073/pnas.2026309118.

Zhang, D., Xia, H., Chen, F., Li, B., Slon, V., Cheng, T., Yang, R., Jacobs, Z., Dai, Q., Massilani, D., et al. (2020). Denisovan DNA in late Pleistocene sediments from Baishiya Karst cave on the Tibetan plateau. Science *370*, 584–587. https://doi.org/10.1126/science.abb6320.