

# Ancient Human DNA: Phylogenetic Applications

**Michael Hofreiter**, *Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

**Linda Vigilant**, *Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

Studies of ancient human DNA attempt to shed light on the genetic relationships among individuals representing archaic or extinct human populations and between these individuals and modern humans. These studies are limited by several factors, including the contamination of ancient samples with modern human DNA, the reliance of necessity upon a single genetic locus, the mitochondrial DNA (mtDNA) molecule, and difficulties in interpretation of results, due to the fact that modern, typically geographically defined human populations are not associated with diagnostic mtDNA sequences.

## Intermediate

### Article contents

- Introduction
- Ancient DNA
- Conclusion and Future Prospects

## Introduction

- 143.1 Ancient deoxyribonucleic acid (DNA) research encompasses the extraction, amplification, and analysis of DNA from samples which are between some decades (e.g., museum specimens) and 100 000 years (Pleistocene animal and human remains, mostly from permafrost or cave sites) old. The materials used are mainly bones or teeth, but sometimes soft tissue such as brain and, more recently, coprolites (Poinar *et al.*, 2001) are also used. Invariably, the DNA recovered from such remains is degraded, although the extent of damage depends upon both storage conditions and age of the sample. Studies of ancient DNA are usually distinguished from forensic DNA analysis and research that uses DNA from noninvasive samples. However, the distinction is not always straightforward, as the latter two fields of research also encounter problems of DNA degradation and sometimes, especially in the case of forensic studies, may even be classified as ancient DNA research.

## Ancient DNA

### Techniques

- 143.2 In principle, the techniques used in ancient DNA research are the same as those used in modern DNA studies, i.e., DNA extraction, amplification by PCR, and subsequent DNA sequencing or, more rarely (and not recommended), restriction fragment length polymorphism (RFLP) or microsatellite typing. However, each of these steps is complicated by the fact that most archaeological remains contain little, if any, endogenous DNA (Handt *et al.*, 1996) and at the same time are often contaminated by modern human DNA (Kolman

and Tuross, 2000). Thus, a number of techniques for sample decontamination prior to extraction have been suggested, including washing or soaking with water or some DNA-degrading chemical (e.g., diluted hydrochloric acid or bleach, not recommended), as well as mechanical removal of the surface. Unfortunately, these methods are only partially effective (Kolman and Tuross, 2000). DNA extraction methods vary, but usually include a decalcification step as well as enzymatic digestion of the sample by protease-K. The DNA is concentrated and separated from polymerase chain reaction (PCR)-inhibiting substances by alcohol precipitation, membrane filtration, or binding to some kind of resin, e.g., silica (Handt *et al.*, 1996; Kolman and Tuross, 2000). Amplification of ancient DNA is done using standard PCR techniques, but with the inclusion of numerous extraction and amplification negative controls and the use of a large number of amplification cycles. The PCR products are analyzed by sequencing, RFLP analysis or microsatellite typing either directly or after cloning of the product. The most reliable results are obtained by the cloning of amplification products and the sequencing of multiple clones (Kolman and Tuross, 2000), while, due to the lack of amplifiable, single-copy nuclear DNA in most ancient human remains (Stone and Stoneking, 1999), it is not usually possible to do microsatellite typing. Both RFLP analysis and direct sequencing produce results in which it is difficult to discriminate between endogenous ancient DNA and modern contamination.

Analysis of ancient DNA is characterized by both a high failure rate and the presence of contaminating DNA. Two techniques have been developed to exclude samples which are unlikely to yield reliable results. The first of these methods is an assessment of the state of biochemical preservation of a sample by amino acid analyses. If either reduction in the total amount of

amino acids or racemization of aspartic acid beyond certain critical values has occurred, DNA retrieval from a sample is highly unlikely (Poinar *et al.*, 1996). Unfortunately, the converse is not true, and a good biochemical state of preservation is a guarantee neither for the presence of endogenous ancient DNA nor for the absence of contamination (Stone and Stoneking, 1999; Kolman and Tuross, 2000). Another method employs quantitation (by competitive or real-time PCR) of starting template molecules for the PCR. When the number of template molecules is low (less than 1000), retrieval of reliable sequences is highly unlikely (Handt *et al.*, 1996), while a sufficient number of starting molecules is by no means a guarantee for the absence of contamination. A complete overview of the recovery and analysis of ancient DNA can be found elsewhere. (See A0342.)

### Phylogenetic applications and pitfalls

143.4 Potential phylogenetic analyses using ancient human samples are hindered by the low quantity, low quality, and almost exclusively mitochondrial origins of ancient DNA. Since one careful investigation has found that only 15% of the samples from an Amerindian cemetery that is only 700 years old yielded amplifiable, single-copy nuclear DNA (Stone and Stoneking, 1999), published analyses of ancient nuclear DNA must be carefully scrutinized to check for adherence to the criteria for reliable analysis of ancient human DNA (Handt *et al.*, 1996; Kolman and Tuross, 2000). Unfortunately, even when reliable mitochondrial DNA (mtDNA) sequences are obtained from ancient samples, the analysis may not yield substantial insights, owing to the fact that some contemporary human mitochondrial control region sequences are shared among individuals of very different geographic origin, so that a person from Sudan may carry the same sequence as a person from Germany. For example, the mtDNA sequence obtained from the 5000-year-old Tyrolean Ice Man 'Ötzi' was identical to that found in several contemporary European individuals, but not in Amerindians. This result supported the authenticity of the specimen by refuting claims that it was a South American mummy brought to the Alps, but revealed little more (Handt *et al.*, 1994). Thus, studies on ancient human populations have to rely on comparisons of the frequency of the various sequence haplotypes or haplogroups found (e.g., Izagirre and de la Rúa, 1999; Stone and Stoneking, 1999; Wang *et al.*, 2000). To effectively do this, however, large numbers of samples need to be analyzed, as even studies of modern populations can reveal strikingly different frequencies. The proper analysis of a large number of ancient human samples is expensive and

time-consuming, and therefore many studies do not fulfil the criteria that are widely accepted as necessary for producing reliable results from ancient human remains (Handt *et al.*, 1996; Krings *et al.*, 1997; Kolman and Tuross, 2000).

Some insights, however, have been provided by 143.5 analysis of ancient DNA. For example, a large-scale study of more than 100 individuals from a 700-year-old Amerindian burial site found the same haplogroups as are present in extant Amerindian populations, suggesting that the population bottleneck in American Indians following European contact did not have a major impact on the genetic makeup of these populations (Stone and Stoneking, 1999). This result is consistent with studies of contemporary American Indians that find a genetic diversity not significantly different from that found in other regions of the world. Population studies may yield interesting results in cases where there are indications from archaeology that a population replacement has taken place. One such study compared present-day, 2000-year-old, and 2500-year-old populations from the same location in China (Wang *et al.*, 2000). The authors found considerable differences between the populations and attributed this to profound changes in population composition over time. As a number of controls were not included, it remains to be seen whether the conclusions will be substantiated or will be found to be artifacts due to contamination.

Another question for which ancient DNA could 143.6 potentially make a significant contribution is in estimating the extent to which the diffusion of Neolithic farmers into Europe beginning some 10 000 years ago affected the genetic composition of the resident preagricultural Paleolithic and Mesolithic populations. One study employed careful precautions to ensure that the reported sequences were indeed endogenous DNA from the ancient bone samples, but presented results from only three samples (Di Benedetto *et al.*, 2000). The two samples, dated between 5000 and 6000 years old, carried sequences still present in Europe, Asia, and Africa today, whereas the 14 000-year-old Mesolithic sample carried a sequence that is unique, but differs by only one nucleotide substitution from some modern sequences in the literature. Adding to this dataset would require immense effort and yield uncertain rewards, particularly as the sequences which are expected from the ancient samples do not differ strikingly from the sequences that could be inadvertently detected from researchers or museum curators of predominantly European ancestry. Taken together, studies on ancient human populations have had relatively little impact on our understanding of human history.

The most important (and probably the only) 143.7 exceptions to the above statement are the molecular

studies on Neanderthal remains (e.g., Krings *et al.*, 1997). Neanderthals lived in Europe from about 300 000 years ago to about 30 000 years ago, although the date of extinction remains controversial. Similarly, there is heated debate on the amount, if any, of genetic contribution by Neanderthals to the modern European gene pool. Analysis of a total of 380 bp of a highly variable portion of the mtDNA produced a sequence bearing considerable differences relative to the sequences of a worldwide sample of modern humans (Krings *et al.*, 1997). Importantly, in a phylogenetic analysis, the Neanderthal sequence fell outside the variation of modern humans. As further evidence against a contribution of Neanderthal mtDNA to the modern human mtDNA gene pool, the Neanderthal sequence was not closer to the sequences from Europe than to sequences found in other regions of the world. These results were later corroborated by additional sequence analyses from the same sample as well as by analysis of two more samples from different geographical regions. Similar significance has recently been attributed to the analysis of ancient DNA from a 60 000-year-old human bone from Australia (Adcock *et al.*, 2001). The authors of this study found a mtDNA sequence which, like the Neanderthal, fell in their analysis outside the variation of all modern humans. Since in all other studies of modern humans the sequences at the base of the human mtDNA tree originate exclusively in individuals from Africa, the finding of a basal sequence in Australia would have important implications for our view of the human population history. However, this study has been criticized on a number of aspects, including the lack of measurement of biochemical preservation, lack of reproduction in a second laboratory, and the reliance on direct sequencing rather than sequences from multiple clones. Finally, the phylogenetic analysis has been questioned, as it is only under certain parameters that the described sequence falls outside the variation of modern humans. Thus, the results of this study need to be corroborated by additional and independent data before they can be seen as a contribution to the understanding of the population history of modern humans. (See A0343.)

## Conclusion and Future Prospects

143.8 The potential for phylogenetic applications of ancient human DNA is clearly limited. With the exception of the research on Neanderthals, few studies have provided new insights into the population history of modern humans. The analysis of Neanderthal DNA has, for technical reasons that are unlikely to change in the near future, been limited to mtDNA. While a contribution of Neanderthals to the mitochondrial

gene pool of modern humans is highly unlikely given the results from the three samples analyzed so far, for nuclear genes, gene flow may or may not have taken place between Neanderthals and archaic modern humans (Cro-Magnons) in Europe. As many nuclear genes have coalescence times older than 600 000 years (the time of the most recent common ancestor of Neanderthals and modern humans; Krings *et al.*, 1997) in modern humans alone, if nuclear sequences are ever obtained from Neanderthals it will often be impossible to distinguish Neanderthal and modern human alleles. More likely in the near future is the analysis of additional mtDNA sequences from Neanderthal remains in order to provide insights about the relative amount of genetic diversity in Neanderthals as compared to modern humans, as well as about the amount of genetic differences between Neanderthal populations from different locations or times. Examination of mtDNA variation of European Cro-Magnon samples up to 40 000 years old is likely to produce sequences of uncertain authenticity and informativeness, since they can be expected to be similar to those from modern humans. Human remains for which the problem of sequence similarity should be less severe could be from the time of emergence of modern humans from Africa approximately 100 000 years ago. Unfortunately, most such remains are from areas with a climate unfavorable for DNA preservation. Moreover, the age of these samples is also at the limit both for the theoretical survival of DNA and for the age of samples from which DNA could so far be extracted successfully. In fact, biochemical analyses have shown that no amino acids and thus most likely no DNA has survived in a number of these fossils (D. Serre, personal communication). The same is true for remains of *Homo erectus* from Asia, the second place besides Europe for which genetic continuity between *Homo sapiens* and an earlier hominid species has been claimed. Thus, it seems likely that, apart from the Neanderthals, the only area where progress can be expected in the near future lies in the execution of painstaking studies on relatively recent, potential population replacements such as during the Neolithic revolution in Europe. Finally, it should be noted that the analysis of ancient animal remains has been extremely fruitful and provides many interesting insights into the phylogeny, phylogeography, and ecology of both extinct and extant species (see the Further Reading section for reviews on these topics).

## References

- Adcock GJ, Dennis ES, Eastaer S, *et al.* (2001) Mitochondrial DNA sequences in ancient Australians: implications for modern human origins. *Proceedings of the National Academy of Science of the USA* **98**: 537–542.

- Di Benedetto G, Nasidze IS, Stenico M, *et al.* (2000) Mitochondrial DNA sequences in prehistoric remains from the Alps. *European Journal of Human Genetics* **8**: 669–677.
- Handt O, Richards M, Trommsdorff M, *et al.* (1994) Molecular genetic analyses of the Tyrolean Ice Man. *Science* **264**: 1775–1778.
- Handt O, Krings M, Ward RH and Pääbo S (1996) The retrieval of ancient human DNA sequences. *American Journal of Human Genetics* **59**: 368–376.
- Izagirre N and de la Rúa (1999) An mtDNA analysis in ancient Basque populations: implications for haplogroup V as a marker for a major paleolithic expansion from southwestern Europe. *American Journal of Human Genetics* **65**: 199–207.
- Krings M, Stone A, Schmitz RW *et al.* (1997) Neanderthal DNA sequences and the origin of modern humans. *Cell* **90**: 19–30.
- Kolman CJ and Tuross N (2000) Ancient DNA analysis of human populations. *American Journal of Physical Anthropology* **111**: 5–23.
- Poinar HN, Höss M, Bada JL and Pääbo S (1996) Amino acid racemization and the preservation of ancient DNA. *Science* **272**: 864–866.
- Poinar HN, Kuch M, Sobolik KD, *et al.* (2001) A molecular analysis of dietary diversity for three archaic Native Americans. *Proceedings of the National Academy of Sciences of the USA* **98**: 4317–4322.
- Stone A and Stoneking M (1999) Analysis of ancient DNA from a prehistoric Amerindian cemetery. *Philosophical Transactions of the Royal Society of London B* **354**: 153–159.
- Wang L, Oota H, Saitou N *et al.* (2000) Genetic structure of a 2500-year-old human population and its spatiotemporal changes. *Molecular Biology and Evolution* **17**: 1396–1400.

## Further Reading

- Geigl EM (2001) Inadequate use of molecular hybridisation to analyse DNA in Neanderthal fossils. *American Journal of Human Genetics* **68**: 287–290.
- Hofreiter M, Poinar HN, Serre D, Kuch M and Pääbo S (2001) Ancient DNA. *Nature Reviews Genetics* **2**: 353–359.
- Kaestle FA and Smith DG (2001) Ancient mitochondrial DNA evidence for prehistoric population movement: the Numic expansion. *American Journal of Physical Anthropology* **115**: 1–12.
- Krings M, Capelli C, Tschentscher F, *et al.* (2000) A view of Neanderthal genetic diversity. *Nature Genetics* **26**: 144–146.
- Kumar SS, Nasidze I, Walimbe SR and Stoneking M (2000) Discouraging prospects for ancient DNA from India. *American Journal of Physical Anthropology* **113**: 129–133.
- O'Rourke DH, Hayes MG and Carlyle SW (2000) Spatial and temporal stability of mtDNA haplogroup frequencies in native North America. *Human Biology* **72**: 15–34.
- Richards M, Cölte-Real H, Forster P, *et al.* (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. *American Journal of Human Genetics* **59**: 185–203.
- Wayne RK, Leonard JA and Cooper A (1999) Full of sound and fury: The recent history of ancient DNA. *Annual Reviews in Ecology and Systematics* **30**: 457–477.

## Glossary

- Bottleneck.** A severe reduction in population size of a species or a population, causing altered gene frequencies. Severe and especially severe and long-lasting bottlenecks considerably reduce the genetic diversity of a population.
- Gene flow.** The exchange of genes between different populations of the same species produced by migration of individuals.
- Gene pool.** Total genetic information present in the reproductive members of a population of sexually reproducing organisms.
- Microsatellites.** Genetic loci containing short (2- to 4-nucleotide) DNA motifs which are repeated several times. The number of repeats can vary greatly among individuals in a population.
- Noninvasive samples.** Samples which can be collected from wild animals without disturbance, e.g., feces, shed hair, shed feathers.
- Polymerase chain reaction (PCR).** An *in vitro* method to multiply a targeted DNA sequence. Use of PCR can produce billions of copies of a certain DNA sequence from as little as a single starting molecule.

**Pleistocene.** Climatic period between 2 million and 10 000 years before present. The Pleistocene was characterized by extreme climatic oscillations, with cold glaciation periods when large parts of Europe and North America were covered by glaciers, interspersed with interglacials, when it was often warmer than today.

**Racemization.** All amino acids (except glycine) exist in two forms, which are mirror images of each other, the D and the L form. In living eukaryotes, only the L form exists. After death of an organism, the L form is converted to the D form until both exist in equal amounts in a dynamic equilibrium. This process is called racemization, and the extent of racemization can be used to assess the state of biochemical preservation of a sample.

**Restriction fragment length polymorphism (RFLP) analysis.** A method that considers variation among individuals in the length of DNA fragments created by an endonuclease.

## Keywords

ancient DNA, human populations, Neanderthal, haplogroup frequency