Proceedings of the British Academy, 77, 267–283

Biomolecular Archaeology: Past, Present and Future

R. E. M. HEDGES[1] & B. C. SYKES[2]
[1]*Research Laboratory for Archaeology, 6 Keble Road, Oxford OX1 3QJ, UK*
[2]*Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK*

Summary. The information potentially available from biomolecules is vast, and often related to the most central concerns of archaeology. It is, however, entirely limited by the survival of such complex molecules and, further, by the sensitivity of the methods for their recovery and analysis. Recent advances in immunology, in protein chemistry and especially in the amplification of DNA sequences are beginning to show that antigens, proteins and nucleic acid sequences can be sufficiently well preserved in archaeological remains for useful results to be obtained. The history of such work shows the importance of establishing approaches that are reliable, specific, and not confused by contamination or diagenesis. Materials such as dried seeds, bone, and mummified skin, as well as artefacts bearing organic traces, have been studied, but a great deal remains to be learnt about the mode of preservation, the degree of biomolecular diagenesis, and the effect of the burial environment. One of the most important results from biomolecular studies is of the genetic constitution of an organism; recent work from several laboratories on the survival of DNA is described, with an account of the archaeological information which can now be contemplated.

Introduction

This review aims to give a unified perspective for a number of different approaches within archaeological science to the biochemical study of ancient

Read 14 February 1991.

remains. Our justification for a wide, rather than more detailed, view is that these approaches are undergoing rapid change, setting exciting promises for the future, and need a broad foundation for their significance to be appreciated.

Biomolecular archaeology can be described as using molecular techniques to obtain biological information about archaeological subjects. It is developing from the interaction of a number of distinct subjects; we here take four separate strands as a foundation. These are, in order of increasing focus:-

i) Which archaeological issues can molecular biology best address?

ii) Which molecules are most worthwhile to study?

iii) How are differences in survival to be taken into account?

iv) Which techniques are appropriate?

Each of these issues influences the others. In this review we consider them in the order given. The treatment must of necessity be brief. In practice, the field is strongly 'technique-led', and this confers a rather unpredictable character to its development.

1. Addressing the archaeological issues

Biological information can refer to an organism's *genetic structure*, or to its *environment*.

Taking the latter first, environmental information can be direct, for example, arising from the estimation of hormones, or of compounds involved in plant defences, or the biochemical effects of unusual or restricted diet. Or it may be indirect, inferred from other (often genetic) information. Examples here might be a malarial environment from a high incidence of the genetically determined condition of sickle-cell anaemia; inference of certain diseases from analysis of antigenically determined immunoglobulins; or, rather generally, the deduction of high population densities and their consequences.

Genetic information itself is rather more definite. It is convenient to distinguish three aspects:-

1.1 Taxonomy (identification)

There is obvious but very useful work to be done in identifying archaeological remains, especially of plants; in analysing relationships between species, e.g., of plants under domestication, or small rodents such as voles; and in the identification of sex of humans and animals. Less obvious possibilities for the future include the identification of diseases, especially viral, microbial and parasitic.

1.2 Phenotypic description

The expression of genetic information (the phenotype) seldom allows individual traits to be associated with the composition of a single gene. Such a situation is most commonly studied for genetic diseases, about which a great deal is now known. Their relevance to archaeology is, however, diminished by the rarity of (serious) genetic disease. One example of a minor defect is that of colour blindness. In the course of time medical and biological research will make the deduction of the phenotype from genetic analysis increasingly possible. Such a deduction could include physical characteristics (e.g., type of hair, colour of eyes), susceptibility to particular diseases, and metabolic characteristics (e.g., the ability for adult humans to digest lactose by secreting lactase; this genetic trait is controlled at a single locus and is associated with the occurrence of a dairy economy).

At present, then, there is little opportunity for deducing the phenotype from information of the genetic composition (although the identification of sex is of course one example). But this aspect can be expected to develop prodigiously over the next decade.

1.3 The genotype and polymorphisms

The timescale of evolution is such that archaeological genetics are mainly concerned with variation within a species, rather than speciation itself. But no mention of speciation should pass up the opportunity to consider the possibility of obtaining information on the genotype of Neanderthal humans. Results so far on archaeological material are encouraging for such speculation. Also, the effect of the severe selection pressure in human plant and animal breeding programmes should enable such cultural events as the neolithic transition to be reflected in major changes in the genotypes of whole populations.

However, as a rule the description of populations in terms of gene *frequencies* is likely to be most relevant. This implies a sufficient sample size to define the particular polymorphic structure. (This issue will be returned to later). A population can be delimited in all sorts of ways:—family groups, geographically isolated communities, cultural (e.g., linguistic) units—and polymorphic variation at the appropriate level must be studied. Three very broad levels are currently useful. These are; '*hypervariable regions*'—where so many alleles are to be found for a series of unlinked loci that only individuals closely related by descent share the same set of alleles (this is essentially the same as genetic "fingerprinting"); '*clinal variation*'—where small differences in the frequencies of alleles at many loci (10—30) can be precisely determined and mapped; and '*population markers*'—where extreme differences have been found to exist in the frequency of a particular allele in a given population

relative to that for neighbouring populations. The two extremes are likely to be most useful in archaeological application. In addition, the analysis of gene frequencies enables the diversity or heterozygosity of a population to be estimated, from which it is possible to infer past effects of the size of the population (hence the recognition of 'bottlenecks').

2. The molecules most suitable to study

The main considerations here are occurrence, stability, and the inherent information content. It is convenient to consider two classes, the linear molecules (comprising proteins and nucleic acids), and non-linear molecules (the rest).

2.1 Non-linear molecular species

Many of these are small, specialised molecules, such as sterols, or polyphenols, which are well-known in natural product chemistry. The class includes more complex types, such as glycoproteins, or polysaccharides (which include the A,B,O blood group antigens). Also, many proteins, although linear in primary structure, are recognised by their secondary or tertiary structure.

Such molecules have not been extensively studied so far, because their chemistry is complicated by diagenetic processes, their information content is often low, and it is very hard to relate what is analysed in the laboratory to what was originally in the organism. But as understanding of diagenesis grows, and laboratory techniques improve, this area, particularly for the lower molecular weights, may well prove to be very worthwhile.

2.2 Linear molecules

Although not all molecules may be worth analysing, (for example collagen, probably the most abundant protein in archaeology, shows very little variation between individuals or species), linear molecules have enormous potential information content. What is more important is that it is much easier to validate the information in such a molecule through sequencing its components (amino acids or bases). Diagenetic change, though serious, can be corrected. Occurrence and stability can vary greatly.

3. What material survives

Understanding the survival of molecular species is crucial to the whole

enterprise. The ability to recover information configured in macromolecules is diminished by the following processes:-

i) Loss of molecular abundance. The decrease of any surviving organic material with time is inevitable. Much of it is leached away through interaction with groundwater movement.

ii) Decomposition of molecular integrity. Even if the majority of atoms remain, chemical changes within the molecules are liable to disrupt the configuration.

iii) Addition of external material. That is, contamination from the environment. This may be as a mixture, or more seriously, contamination may combine chemically with the chosen molecule (such as the cross linking between polyphenols and protein).

Survival is dependent on the particular environment, the type of tissue, and the molecular species. Occasionally survival is spectacular, and environments may be ranked in terms of the survival of their material.

3.1 Environment

As examples:

i) A fossil Miocene *Magnolia* leaf, has been shown to contain recoverable DNA from the Clarkia deposits (Golenberg *et al.* 1990).

ii) The Windover site (Doran *et al.* 1986; Lawlor *et al.* 1991) in Florida, U.S.A. contains nearly 100 well preserved skulls containing soft tissue from 8000 BP. (The skeletal material is buried in neutral waterlogged peat).

iii) Frozen mummies from Peru, mammoths from the U.S.S.R. (Guthrie 1990) and more recent human burials from Greenland (Hart Hansen *et al.* 1991) have been studied.

iv) The deliberately mummified material from Egypt (Pääbo 1985).

v) Bog bodies (e.g., Lindow Man; Stead *et al.* 1986).

vi) Well-preserved bone; e.g., from the "Mary Rose" shipwreck of AD 1545.

vii) Bone and teeth from N.W. Europe—commonly found with several percent of total organic material remaining.

viii) Bone from hotter climates, where frequently there is less than 1 % of original protein remaining.

Obvious generalisations suggest that low temperature, low oxidation, low water activity (i.e., minimal wetting/drying cycles or water flow regimes), and

above all, suppressed microbiological activity, are conditions for improved survival. At present our understanding of survival is more anecdotal than systematic, however.

Time, itself, has not been explicitly invoked as a determinant of survival. Certainly we would expect 'older' material containing biomolecular information to come from a far narrower range of environments. Very little organic study has been made on bone older than the Last Glacial Maximum, and it is too soon to predict with any confidence the maximum age for which hominid remains, for example, will be susceptible to biomolecular study.

3.2 Tissues

Early work concentrated on preserved soft tissue, partly because the abundance of DNA and protein is in general higher, but mainly because conditions which preserved the tissue as a whole might be expected to be conducive to preserving macromolecular integrity also. Although preserved soft tissue is not common, enough examples abound to enable plenty of investigations to be made. However, there is no doubt that the possibility of extracting information from commonly preserved tissues such as bone, tooth, and perhaps carbonised seed and pollen, dramatically increases the value of the whole approach.

Survival in a given tissue will be affected by microbiological action, by the local chemistry, and to an extent mechanically. Although microbiological action is ubiquitous, it is likely to diminish in time for bone, as limited resources are consumed, and it is interesting that a recent study (Thuesen and Engberg 1990) identified *Acetomycetes* contamination in preserved soft tissue but not in the associated bone. Bone has another point of interest, which is the affinity between hydroxyapatite and both protein and DNA. Hydroxyapatite columns have been used in the purification of DNA (Bernardi 1969), and may help to retain DNA and stabilise it from chemical degradation. (It appears that such is the case in preserving osteocalcin in situations where nearly all the collagen is lost). Bone can also act to buffer the pH of the local aqueous environment.

At this stage virtually nothing can be said about possible survival in pollen or carbonised seed (where the temperature of carbonisation is obviously crucial). Although not a tissue, the survival of proteins in 'residues' on stone tools (Loy 1983) has been claimed. The evidence for this is persistent, and certainly warrants fuller investigation than it has so far received. Of course the number of samples is not so very large, and replication of results is thereby made difficult.

3.3 Molecular species

It is well known that lipids survive in a wide range of Quaternary environments (Cranwell 1981), although they also undergo much molecular modification during burial. Polyphenols too are known to be stable, often cross-linking to form macromolecular aggregates, and contributing to the stability of humic and fulvic acids. A major problem with blood group determination (ABO system) is that the difference between the A and B antigens (oligosaccharides) is chemically very slight, and conversion may easily occur through chemical or biological action during burial.

Probably proteins and DNA need some form of 'protection', if only to reduce the rate of dissolution by groundwater movement. As mentioned above, different proteins appear to survive differentially, depending upon their immediate chemical environment (such as hydroxyapatite). The survival of RNA, generally considered to be less stable chemically than DNA, has not been systematically investigated. However, predicting the survival of specific molecules cannot be relied upon—few would have expected collagen to be so much more completely lost from bone than many of the plasma proteins; or for DNA to survive in a Miocene fossil under almost any conditions—and therefore any molecule for which reliable analytical techniques exist might reasonably be investigated.

A further aspect of molecular survival is the information it might give concerning the environment. For example, amino acid residues in protein that are easily oxidized (serine; methionine; cysteine) may find use as an environmental indicator, which in turn might predict the survival of other species.

4. Techniques

The low abundance, degradation and liability to contamination of the molecular species being analysed demand that techniques be extremely sensitive, highly specific, and can operate with poorly characterised or impure material.

4.1 "Conventional" chemical methods

These include standard separation methods, principally high pressure liquid chromatography and gel electrophoresis. Such methods are adapted to handle nanogram-picogram quantities of material. Major problems arise if one is dealing with material degraded to the point that it presents a continuum distribution, since no meaningful separation can then be effected. In principle polypeptides can be sequenced (e.g., by automatic sequencers using Edman degradation methods). so long as they have been completely purified

first. Mass spectrometric methods are very sensitive and can operate with mixtures, and large molecules (e.g., polypeptides) can also now be analysed. While chemical methods are necessary in any case to bring about a concentration of the molecule under study, they are most useful in the study of smaller molecules (lipids; polyphenols; oligosaccharides) which degrade to a more discrete mixture than do protein and nucleic acids.

4.2 Immunochemical methods

The remarkable sensitivity and specificity of the antibody-antigen reaction enables, in principle, proteins and similar antigens to be detected if the relevant antiserum can be obtained. Early work suffered from both false positives and false negatives in detection because neither the specificity of the antiserum nor the integrity of the antigen could be adequately guaranteed. Technical improvements in sensitivity include particularly enzyme-linked inhibition assays (ELISA) (see, for example, Smith and Wilson 1990), and improvements in specificity have been brought about with the use of monoclonal antibodies (for an example, see Cattaneo *et al.* 1990). It is likely that the more specific the method, the smaller the detectable fraction of surviving biomolecule. This is because diagenesis is likely to reduce greatly the chance of a molecule retaining its epitopes intact. Therefore, in many archaeological applications, it seems that some form of compromise between specificity and sensitivity will be necessary if the changes in molecular configuration are to be studied.

4.3 DNA methods

The chemistry of DNA is frequently able to exploit its unique linear stranded complementary structure. Two approaches in particular that manage this are hybridisation and the polymerase chain reaction.

In hybridisation, the affinity between a DNA probe of known composition and the test DNA can be sensitively estimated (at the level of nanograms) in a manner analogous to immunochemical reaction. Here the probe binds to a complementary DNA if it is sufficiently homologous (about 90%).

In the "polymerase chain reaction" (PCR) an enzymic system is used to "amplify" (increase the number of) a template DNA sequence by repeated replication. After 30–40 repeats an amplification of over a million can be achieved. The template DNA sequence is typically 100—1000 bases long, and is defined by the choice of "primers" which are oligonucleotides of about 20 bases in length, complimentary to the beginning and end of the template sequence. Since the PCR can function in the presence of overwhelming quantities of (degraded and foreign) DNA, it is able to be both specific and

extremely sensitive in that the chosen sequence (template) can be amplified against the background of all the impurities to a level where further analysis (such as sequencing) is possible. The PCR method is relatively new, and still undergoing rapid development. Technically, there are many difficulties in its application (for example, the amplification of contamination, non-specific amplification, inhibition of the enzymic system by unknown impurities), but the combination of sensitivity and specificity has enabled it to become the major technique in understanding ancient DNA (Pääbo *et al.* 1989).

5. Results

The results summarised here make no attempt to be comprehensive, but to point out some of the outstanding recent work. We believe that at least most of the published work on archaeological DNA sequences has been included.

5.1 Molecules other than proteins and nucleic acids

The analysis and identification of lipids is considered in this volume (Evershed *et al.* this volume; though with reference to a particular environment), and such compounds form the subject of a great deal of investigation in sediments (Cranwell 1981; Brassell 1985). Little work has been done so far on specific identification of lignins and polyphenols, at least in terms of their biological significance. However, the field is very broad, and much work has been published covering various aspects from the diagenesis of amber, the identification of resins on stone tools, to the ageing of varnishes in painting.

5.2 Proteins

We mainly consider proteins shown to survive in bone. Other materials have been studied, for example shell (Curry *et al.* in press) and "residues" on stone tools (Loy 1983), as well as preserved soft tissue, but bone is both the most studied and the most liable to provide biologically significant information.

The survival of collagen (90% of bone protein) is well known but is not very genetically informative. However, in environments where collagen is almost entirely lost, it appears that other non-collagenous proteins may have survived very much better. In particular osteocalcin has been extracted in large enough quantities for radiocarbon dating (Ajie *et al.* 1990). The immunological titre appears to be about 1 % of the extractable protein. Most of the other proteins recorded in bone in recent publications are plasma proteins. They include haemoglobin (identified by ELISA) (Smith and Wilson 1990), albumin (using monoclonal antibodies) (Cattaneo *et al.* 1990) and the immunoglobulins (using electrophoresis and immunoblotting)

(Tuross 1988). To our knowledge, no-one has yet detected a protein polymorphism, but the steady development of technique and the fact that several groups now have good evidence for the survival of non-collagenous proteins in bone over more than 2000 years is very promising.

5.3 Nucleic acids

Firstly, it is not yet clear to what extent *RNA* may survive. Being chemically much less stable, it is unlikely to be as worthwhile to study as DNA.

DNA has been studied in preserved soft tissue, in bone, and in dried and partially carbonised seeds. There is now reasonably good evidence that DNA survives consistently enough in bone (Hagelberg *et al.* 1989; Horai *et al.* 1989; Hanni *et al.* 1990; Thuesen and Engberg 1990; Williams *et al.* 1990) that its study in this tissue should be taken very seriously indeed, since bone is the most ubiquitous of archaeologically important tissues.

Most tissues appear to yield of the order of micrograms of DNA extracted from about a gram. If this is analysed for molecular weight by gel electrophoresis, it is shown to have a range of low molecular weight components (< 1000 bp in length), as well as high molecular weight material. All of this DNA may be non-indigenous, and in any case only a very small fraction appears to be chemically intact. Failure to detect DNA at this stage may not preclude failure to detect specific DNA later on. Pääbo in particular has analysed the chemical integrity of this extracted DNA (Pääbo 1989) and shown it to have extensive damage (for example, very few purine bases remained, and of the DNA that could be analysed, at least 5% of all sites were damaged). At this stage of investigation it seems likely that there is a correlation between apparent DNA preservation and apparent protein preservation—this would after all be expected—but in general still too little is known about conditions and correlates favouring DNA survival.

Some success has been achieved through hybridisation experiments. In these, purportedly human DNA has been hybridised with human DNA probes, mainly using mtDNA probes, but also repeat copy probes such as Alu (Thuesen and Engberg 1990; Williams *et al.* 1990). These show that perhaps one part in 10^5 of the extracted DNA is in sufficient condition to hybridise. It is still too early to be clear how reliable and specific the hybridisation approach can be; it has the advantage of being relatively straightforward and is less subject to contamination problems than is the PCR approach. The information achieved from hybridisation is essentially to learn of the specificity of a particular probe, so that species, or sex, rather than a distinct sequence, would be the main result. However, the technique is likely to remain useful as an overall screening method.

To be sure that indigenous DNA is present it is necessary to recover

sufficient material to demonstrate that it has the 'correct' sequence. This may not be a sufficient criterion, since contaminant DNA may also have the correct sequence—it depends upon what sequence is being compared, and the nature of contamination. The most difficult case, although inevitably the one most chosen so far, is the analysis of human DNA, since contamination by other human DNA is quite feasible. To generate enough DNA for sequencing, some form of cloning or amplification is necessary, and this is now best achieved using the PCR. The number of published reports of DNA detected by PCR amplification from archaeological material is still very limited, and only a small fraction of such work has gone on to sequence the product. Virtually all the successful PCR so far has been on mtDNA, presumably because of its high copy number relative to nuclear DNA. Most of the DNA sequences that have been amplified are shorter than 200 bp, although evidence is beginning to accumulate that DNA from bone may be amplifiable in longer lengths (up to 400 bp). As already mentioned, the use of PCR carries with it a number of difficulties, which are described below.

5.4 Difficulties

5.4.1 Contamination

Contamination of the sample by genomic DNA, before the PCR reaction is carried out, undoubtedly can occur. How serious this is (in the sense of how frequent and how unavoidable) is not yet clear. The most likely contamination is by human DNA, and this has been observed when analysing non-human material. Too few such studies have been made for conclusions to be drawn at this stage. It is possible that modern contamination would amplify much better in longer lengths than ancient DNA, so that its presence might be generally detectable with enough work. Contamination by laboratory-derived DNA sequences is also a very real danger, but this can usually be detected by the careful use of controls. However, it is imperative that laboratories working in this field do take adequate precautions and publish controls, so that there is no room for doubt.

5.4.2 Non-specific amplification and 'jumping' PCR

Some experimental conditions may make the reaction less specific; this will be obvious if the product has a different molecular weight from the template; in any case, it is generally thought wise to sequence or hybridize the product in case of any doubt. A related complication is the ability for the reaction to 'jump' from one replicated strand to the next when DNA strands contain damage (Pääbo *et al.* 1990). This makes it possible for amplification to take

place even when less than one intact template sequence is present. The resultant amplified product will be a mixture, from which a consensus can be obtained by sequencing. The ability to 'jump' is in general very useful, since a higher degree of damage can be sustained whilst still obtaining sequence information. However, it may not be the best way to reach consensus information from damaged sequences, and in any case where several alleles may be present (as in the amplification of diploid genes), it is possible for the PCR to bring about what amounts to recombination between them, and this may limit the information available from such genes.

5.4.3 Inhibition

Several groups have demonstrated that preparations made from extracted ancient DNA will inhibit the PCR. Treatment with a protein (e.g., bovine serum albumin) to precipitate possible contaminants (haem has been suggested) has been shown to help. However, there can be many ways in which the unknown degraded material present in archaeological preparations may disrupt the course of the PCR.

5.4.4 Sequencing

Most PCR products will require sequencing, and so far the method, although standard and well-proven, appears to be rather more difficult on amplified ancient DNA. In any case, sequencing is relatively time-consuming, and should be avoided if other forms of evidence are appropriate. However, if general polymorphic information is sought, it may be difficult to avoid the extra effort.

5.5 Summary of results

From most surviving tissues, and especially bone, enough DNA appears to survive for at least a few thousand years that it is possible to obtain sequence information at least of mitochondrial DNA. There remains a great deal to learn about the best way to operate the PCR, and to learn about the conditions for DNA survival, and the degree of damage to the remaining DNA, so that many more possibilities are likely to emerge with time; for example, the amplification and sequencing of single copy genes from animals. In our view, the most important finding at present is that many bones (very roughly between 10 and 50%) can be analysed for mtDNA. Nothing in the experimental work reported so far gives any indication as to how far back DNA might usefully survive. No doubt suitable material will become a decreasing fraction of the total surviving corpus of bone, itself a decreasing

quantity with time. There is at least reasonable hope that Neanderthals and animals of their time may eventually be subjects for study.

6. Applications and future prospects

Applications to archaeology so far have been almost negligible, mainly because very little significant information can be obtained by a one-off analysis. Perhaps the most important pointer to future work is the analysis of the mtDNA control region sequence from a brain from Windover (Pääbo *et al.* 1988) which turns out to be unique (so far) for North American modern Indians, but similar to sequences common in East Asia. This result immediately points out the necessity to interpret any sequence data set in context, and of course there is as yet no very appropriate context for archaeological material.

In our view, the best way forward both builds from relatively modest archaeological projects which can supply their own context, and also emulates those successful studies of modern human populations which can act as analogues to archaeological populations. In both cases the requisite sample size of necessity should be small (< 100). The analysis of 100 samples by PCR followed by sequencing is by no means out of the question, and will no doubt become easier as the technology is developed, but the availability of samples, especially for pre-Neolithic times, will always be limited.

Two modern studies must stand as examples. The first (Cann *et al.* 1987; Horai and Hayasaka 1990) concerns the restriction fragment analysis (i.e., at less resolution than sequencing) of mtDNA from (initially) 150 humans world-wide. This has provided persuasive evidence for a bottleneck in the population of mitochondrial lineages estimated at 100,000—200,000 years ago, with the further suggestion that a founder population moved out of Africa at that time. Clearly this picture would be enormously enhanced if similar analyses on well-sampled individuals could be made for successive time-slices in the past (even if only as far back as 20,000 years).

Secondly, a recent study on 15 individuals from the !Kung of South Africa, with 68 control individuals from neighbouring regions (Vigilant *et al.* 1989) was able to recover interesting and relevant information on the diversity within and between populations. (In this case, two 400 base pair sequences were analysed in the mtDNA control region.) Two conclusions, especially relevant for archaeology, stand out. One is that the effective breeding size could be estimated as 5000—a figure, not unexpectedly, consistent with the known value—and secondly, that the average amount of migration (females only, since only mtDNA is observed) is not more than 13 metres per year. Again, it is quite possible to imagine a similar study being made on arch-

aeological populations (although here the definition of population requires some careful consideration), leading to measures for the same parameters.

6.1 Archaeological projects

Firstly it is useful to distinguish between those on isolated or a very few individuals, and those made on populations.

On *individuals*, the first uses would be to act as controls, where the 'correct' answers are already known. For example, the assignment of species, and of sex. (Sex can, in principle be assigned through amplification of high copy number sequences found only on the Y chromosome. Several groups informally report achieving this, but none appears to be able to do so consistently at present.) Having established this, the same techniques can be used for sexing unknown individuals (human and animal), and to the identification of species. If DNA survives for long enough, palaeofauna can be studied, so that evolutionary relationships and ecological change can be followed.

For *populations* the analysis of burial assemblages is an obvious choice. In Oxford we are starting work on Anglo-Saxon cemeteries (such as Berinsfield and Lechlade in Oxfordshire), for which there is evidence (Härke 1990) that the possession of weapons is correlated with non-metrical epigenetic traits in the skeletons. It should be possible to relate mtDNA lineages with such cultural, epigenetic and possibly spatial information. Such work can be easily extended to cover the possible differentiation between Romano-British and Anglo-Saxon lineages, and also subsequent invasions (Viking and Norman). It should be noted that the first generation of invading warriors would not contribute their mtDNA lineages to the local community until invading females arrived and became established. A similar study might be made on burial assemblages in neolithic chambered tombs.

In these cases it is assumed that only mtDNA sequences are accessible to study. A reasonably optimistic view is that with time other highly polymorphic genes will be identifiable, and greatly contribute to our knowledge of the genetics of ancient populations. However, at this stage such work lies too far into the future to be clearly seen.

Acknowledgements

Our own work in this field is being supported by a Research Grant from S.E.R.C. We should also like to acknowledge here the contribution by our colleagues Martin Richards and Kate Smalley.

References

Ajie, H.O., Kaplan, I.R., Slota, P.J. and Taylor, R.E. 1990: AMS radiocarbon dating of bone osteocalcin. *Nuclear Instruments and Methods* B52, 433–438.

Bernardi, G. 1969: Chromatography of nucleic acids on hydroxyapatite. *Biochimica et Biophysika Acta* 174, 423–434.

Brassell, S.C. 1985: Molecular changes in sediment lipids as indicators of systematic early diagenesis. *Philosophical Transactions of the Royal Society of London* A315, 57–75.

Cann, R.L., Stoneking, M. and Wilson, A.C. 1987: Mitochondrial DNA and human evolution. *Nature* 325, 31–36.

Cattaneo, C., Gelsthorpe, K., Phillips, P. and Sokol, R.J. 1990: Blood in ancient human bone. *Nature* 347, 339.

Cranwell, P.A. 1981: Diagenesis of free and bound lipids in terrestrial detritus deposited in a lacustrine environment. *Organic Geochemistry* 3, 79–89.

Curry, G.B., Cusack, M., Endo, K., Walton, D and Quinn, R. in press: Intracrystalline molecules from brachiopods. In *Proceedings of the 6th International Conference on Biomineralisation, Odawara, Japan* (Tokyo, Springer-Verlag).

Doran, G.H., Dickel, D.N., Ballinger, W.E., Agee, O.F., Laipis, P.J. and Hauswirth, W.W. 1986: Anatomical cellular and molecular analysis of 8000 yr-old human brain tissue from the Windover archaeological site. *Nature* 323, 803–806.

Golenberg, E.M., Giannasi, D.E., Clegg, M.T., Smiley, C.J., Durbin, M., Henderson, G. and Zurawarski, G. 1990: Chloroplast DNA sequence from a Miocene Magnolia species. *Nature* 344, 656–657.

Guthrie, R.D. 1990: *Frozen Fauna of the Mammoth Steppe: the Story of Blue Babe* (Chicago, University of Chicago Press).

Hagelberg, E., Sykes, B. and Hedges, R. 1989: Ancient bone DNA amplified. *Nature* 342, 485.

Hanni, C., Laudet, V., Sakka, M., Begue, A. and Stehelin, D. 1990: Amplification de fragments d'ADN mitochondrial a partir de dents et d'os humains anciens. *Comptes Rendues d'Acadamie de Science de Paris* 310, 365–370,

Härke, H. 1990: "Warrior Graves"? The background of the Anglo-Saxon weapon burial rite. *Past and Present* 126, 22–43.

Hart Hansen, J.P., Meldgaard, J. and Nordqvist, J. 1991: *The Greenland Mummies* (London, British Museum publications).

Horai, S., Hayasaki, K., Murayama, K., Wate, N., Kioke, H. and Nakai, N. 1989: DNA amplification from ancient human skeletal remains and their sequence analysis. *Proceedings of the Japanese Academy* 65, 229–233.

Horai, S. and Hayasaka, K. 1990: Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *American Journal of Human Genetics* 46, 828–842.

Lawlor, D.A., Dickel, C.D., Hauswirth, W.W. and Parham, P. 1991: Ancient HLA genes from 7,500-year-old archaeological remains. *Nature* 349, 785–788.

Loy, T.H. 1983: Prehistoric blood residues: detection on tool surfaces and identification of species of origin. *Science* 220, 1269–1270.

Pääbo, S. 1985: Preservation of DNA in ancient Egyptian mummies. *Journal of Archaeological Science* 12, 411–417.

Pääbo, S., Gifford, J.A. and Wilson, A.C. 1988: Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Research* 16, 9775–9787.

Pääbo, S. 1989: Ancient DNA: extraction, characterisation, molecular cloning and enzymatic amplification. *Proceedings of the National Academy of Sciences of the USA* 86, 1939–1943.

Pääbo, S, Higuchi, R.G. and Wilson, A.C. 1989: Ancient DNA and the Polymerase Chain Reaction. *Journal of Biological Chemistry* 264, 9709–9712.

Pääbo, S., Irwin, D.M. and Wilson, A.C. 1990: DNA damage promotes jumping between templates during enzymatic amplification. *Journal of Biological Chemistry* 265, 4718–4721.

Smith, P.R. and Wilson, M.T. 1990: Detection of haemoglobin in human skeletal remains by ELISA. *Journal of Archaeological Science* 17, 255–268.

Stead, I.M., Bourke, J.B. and Brothwell, D. (editors) 1986: *Lindow Man: the Body in the Bog.* (London, British Museum Publications).

Thuesen, I. and Engberg, J. 1990: Recovery and analysis of human genetic material from mummified tissue and bone. *Journal of Archaeological Science* 17, 679–689.

Tuross, N. 1988. Recovery of bone and serum proteins from human skeletal tissue: IgG, osteomectin and albumin. In Ortner, D.J. and Aufderheide, A.C. (editors), *Symposium on Human Palaeopathology: Current Syntheses and Future Options* (Washington, Smithsonian Institution).

Vigilant, L., Pennington, R., Harpending, H., Kocher, T.D. and Wilson, A.C. 1989. Mitochondrial DNA sequences in single haris from a southern African population. *Proceedings of the National Academy of Sciences of the USA* 86, 9350–9354.

Williams, S.R., Longmire, J.L. and Beck, L.A. 1990: Human DNA recovery from ancient bone. *Journal of Physical Anthropology* 81, 318.

APPENDIX: Some relevant characteristics of DNA

For those unfamiliar with the molecular biology of DNA we include the following relevant facts:-

i) Most DNA resides in the nucleus of each cell of an organism, usually as a complex with protein. The nucleus of somatic cells in animals contains two copies of the entire DNA sequence, divided between 23 chromosomes, one copy deriving from each parent. Male humans have a Y chromosome inherited from their fathers (males have one X, and females have 2 X chromosomes, inherited from either parent). The total sequence of nuclear DNA amounts to approximately 3×10^9 bp or about 0.03 ng per cell. Animals also contain a small amount of DNA in their mitochondria. Each mitochondrion contains about ten copies of a DNA sequence of about 16,000 bp, and is maternally inherited. There are several hundred mitochondria in most cells, and therefore far more mtDNA sequence copies than nuclear DNA sequence copies.

ii) Some expressed systems are highly polymorphic, notably the MHC gene complex on chromosome 6, where the variation within a population is there to provide an extensive repertoire of immune responses to infection. There are sufficient MHC alleles to be useful as genetic markers for individuals and could be used, for example, to explore kinship relationships in burial groups.

iii) Genetic variation is even higher in DNA sequences that are not expressed in the phenotype since new mutations are not eliminated by selection. Most powerful as genetic markers are the blocks of DNA where alleles

differ in the number and arrangement of tandemly repeated oligonucleotide blocks. These variable number of tandem repeats (VNTR) systems are the basis for genetic fingerprinting and would be immensely valuable in precisely defining individual relationships.

iv) Less valuable but more accessible are the variable regions of mitochondrial DNA (mtDNA). They are less valuable in defining kinships because mtDNA is inherited only through the maternal line so, for instance, brothers, sisters and their maternal aunts and uncles would all be indistinguishable. However, the maternal inheritance and the lack of genetic recombination does allow construction of mitochondrial lineages which will survive intact from ancient to modern times thereby allowing populations to be traced over considerable periods. It is more accessible because there are at least a hundred copies of mtDNA per cell compared to only two for nuclear genes, so when DNA survival is low there is a greater chance of recovery.

v) A feature of DNA sequences of particular relevance to ancient DNA studies is the copy number, which determines the likely concentration of a specific DNA sequence. Although most nuclear genes are single copy, many families of sequences exist which are copied throughout the genome, often to very high copy numbers (e.g., 10^4). A well known example is the Alu family. mtDNA has already been mentioned, with multiple copies by virtue of the large number of mitochondria per cell.

vi) A sequence of only about 20 bp has a chance of 4^{20} (about 1 in 10^{12}) of occurring at random, and so is virtually unique. Most ancient DNA amplifications are of about 100–400 bp in length, and it would be difficult to reconstruct long sequences in this way. Most genes are 100–10,000 in length, and molecular biology techniques are well adapted for such lengths. The length of VNTRs, which must be completely amplified for polymorphic information, are usually too long for present methods, unless the repeating segment is very short.