MEDICINA NEI SECOLI ARTE E SCIENZA, 18/3 (2006) 701-724
Journal of History of Medicine

Articoli/*Articles*

AUTHENTICITY IN ANCIENT DNA STUDIES

M. THOMAS P. GILBERT* AND ESKE WILLERSLEV

Center for Ancient Genetics, Niels Bohr and Biological Institutes,
University of Copenhagen,

SUMMARY

AUTHENTICITY IN ANCIENT DNA STUDIES

Ancient DNA studies represent a powerful tool that can be used to obtain genetic insights into the past. However, despite the publication of large numbers of apparently successful ancient DNA studies, a number of problems exist with the field that are often ignored. Therefore, questions exist as to how reliable the conclusions of many of the published studies are. In this paper we outline first the problems associated with aDNA studies, and secondly present potential guidelines designed so as to enable non-specialist readers the opportunity to critically assess the quality of aDNA publications.

Introduction

The last twenty years have seen the publication of a large number of 'successful' ancient DNA studies, investigating a broad range of topics. Ancient DNA (aDNA) techniques have allowed the analysis of species and populations that no longer exist[1], and provided a means of *directly* tracking genetic changes through time[2]. Unfortunately though, while many aDNA studies are valuable, a skeptic might argue that their sheer number masks from the average reader the fact that the field is riddled with challenges and pitfalls. Some of these are common with other disciplines that study the

Keywords: Ancient DNA – Degradation – Survival – Contamination - Authenticity

material record of the past; problems of provenance, dating and ultimate interpretation. However, aDNA studies are also subject to problems peculiar to the field, which manifest themselves as difficulties in generating sufficient authentic DNA sequences to make a study conclusive. These problems arise as a result of the *post mortem* degradation of DNA, either through the generation of miscoding lesions which can lead to sequence errors, or through the physical destruction of the molecule, increasing the risk of preferentially amplifying a contaminant sequence.

Over the past 15 years, various authenticity criteria have been published in response to the growing awareness of the problems that are associated with the study of ancient DNA[3]. However, although these lists were designed with a specific aim in mind – the provision of researchers with guidelines that would ensure the generation of authentic data – they also suffer from weaknesses. Due to the complications associated with a DNA, they can both hinder the publication of good studies that do not adhere to all the criteria, and also enable the publication of erroneous results that adhere strictly to them. This is due to the fact that these lists present inflexible checkpoints aimed at avoiding or detecting bad data, without requiring the scientist to consider the underlying processes involved. Therefore it has been argued that a more reliable, cognitive approach towards the assessment of the data reliability is required[4].

This article outlines the current knowledge about what can go wrong in a DNA studies, and what solutions may exist, by expanding on a previous article[5] with a more generalist audience in mind. With such information to mind, we believe that even non-specialist readers will be better positioned to determine whether or not the data is real and the results of a study significant. Thus we initially address what the serious problems are within the field, and then describe how reviewers, readers and authors may assess a publication. We hope this will then enable interested parties to weigh the criteria specifically to the problem on a case-by-case basis. Lastly we highlight to the reader that, whilst we focus

predominantly on problems associated with studies of human material, any studies that utilises degraded DNA may encounter similar problems.

What makes an aDNA result authentic?

Four main factors affect the authenticity of aDNA studies. Three of these are associated with the initial generation of the data

(1) The likelihood of any DNA being in the sample

(2) The ability to differentiate contamination from endogenous DNA

(3) The accurate amplification of DNA

While the fourth asks:

(4) Is there enough data to support the conclusions?

We expand on these below.

(1) The likelihood of any DNA being in the sample.

Despite a number of reports that detail methods designed to establish the extent of DNA survival (see below), this is a complex problem, prone to the vagaries of *post mortem* processes. Ideally, the rate of DNA degradation in ancient tissues would correlate with experimentally determined, *in vitro* rates. Unfortunately however, many other factors appear to alter these rates, as can be seen through a brief consideration of the various damage processes that may affect DNA.

Post mortem DNA damage processes

In most scenarios, the degradation of endogenous DNA (*i.e.* that belonging to a sample of interest) commences shortly after the death of the host. In humans, within 4-5 minutes after death cell autolysis initiates[6] . As the cells of the body are deprived of oxygen, carbon dioxide in the blood increases, pH decreases and wastes accumulate to cytotoxic levels. Concomitantly, unchecked cellular enzymes, including lipases, proteases, amylases and nucleases, begin to dissolve the cell from the inside out. Soon the cells rupture, releasing nutrient-rich fluids that encourage the growth of internal and environmental micro-organisms. As they spread through the

corpse, these organisms will contribute to the degradation of DNA both directly, by nuclease attack and indirectly, by breaking down the bone matrix that the DNA is lodged within through dissolution of the mineral phase and protease-digestion of the organic phase[7].

Therefore, despite the fact that most human diploid cells contain several billion bases of nuclear DNA, and thousands of copies of mitochondrial DNA, its decay may be so fast that within months, if not weeks, no PCR-amplifiable template remains[8] (PCR, the Polymerase Chain Reaction is the most common tool used to investigate aDNA. In brief it involves the cyclical replication of specific DNA sequences of interest, amplifying exponentially the original target up to exceedingly high numbers. This is undertaken to simplify subsequent genetic sequence analyses).

In some cases, these degradation processes may be arrested, and the host DNA will become stabilized for longer periods. During this time, slower acting chemical processes of DNA degradation will modify the molecule. These biochemical modifications are believed to be analogous to those seen *in vivo*, and involve both oxidation and hydrolysis of the molecule, resulting in crosslinking and fragmentation of the molecule's chemical backbone and the alteration of individual nucleotide bases[9].

DNA decay is thus caused by a range of chemical processes, and therefore the rate at which decay occurs is dependent upon the factors that control any chemical reaction: temperature, the concentration of reagents, extent to which those reagents can interact, and the availability of co-factors and competitor molecules. In practice, this means that degradation of DNA may be significantly reduced at low temperature, in the absence of chemically free water (by freezing or drying), where pH is kept above neutral and where microbial presence is limited by procedures which curtail colonization. This latter case may explain why butchered bones have significantly better histological preservation than unbutchered ones[10].

Although the exact contributions of individual processes to the damage will vary with the direct environment surrounding

specimens these processes are ubiquitous enough to suggest that even in optimal environments for survival (*e.g.* buried deep in permafrost), DNA is unlikely to survive for more than a million years[11].

In general therefore, samples that have been kept cold and desiccated (including by the presence of high salt levels) have been observed as good sources of aDNA[12]. Similar observations have also been made on samples extracted from anaerobic environments[13]. Furthermore, it seems that samples with well-preserved microstructure may yield superior DNA, presumably because they prevent the entrance of water and other molecules[14]. Lastly interactions between DNA and other chemicals within a sample may modify the rates of DNA degradation. One demonstrated example that appears to retard degradation is the binding of DNA to hydroxyapatite (a major component of bone)[15].

Estimation of DNA survival and quality

Based on what is known about DNA and other biomolecular degradation, various analytical methods have been proposed to help estimate DNA survival in target samples[16]. The most frequently advocated method is to measure the racemization of the L to D - enantiomers of aspartic acid residues, proposed to occur at a rate similar to that of DNA depurination[17], in order to gauge whether DNA also survives in a particular specimen. An alternative approach has been to use gas chromatography/mass spectroscopy (GC/MS) to measure levels of other biomolecules in samples that also correlate with DNA survival. One example is to assay levels of oxidized pyrimidines 5-hydroxyuracil and 5-hydroxy-5-methylhydantoin, which inversely correlated with DNA retrieval[18]. It has also been demonstrated that samples that contain reproducible DNA sequences yield abundant pyrolysis products assigned to 2,5-diketopiperazines of proline-containing dipeptides[19]. A third method, not reliant on biomolecular proxies, is to calculate a sample's thermal age (the number of years required, at a constant 10°C, to produce the degradation calculated from the samples

thermal history)[20] and derive the expected amount of DNA depurination under these conditions. Although thermal age may give some idea about the general preservation at a site it should be kept in mind that DNA preservation varies tremendously even between samples of the same age from a single site. Regardless of which tests are used, we caution the reader that at present all techniques involve correlations that are based on limited studies, assay factors which appear to vary a great deal with the environment, and thus cannot be expected to provide generally applicable estimates of DNA survival. Therefore it may be sensible not to use them to predict DNA survival, but rather, to confirming that sequences already retrieved may be authentic. Thus although useful, it is essential to appreciate that the above methods cannot prove or disprove the authenticity of aDNA studies. However, where the extreme ages and environmental conditions clearly argue against the survival of any useful DNA[21], the investigators should be able to propose a plausible mechanism for DNA survival.

While good preservation can help suggest DNA survival, it provides little information on the presence or absence of contamination within an ancient sample.

2. The presence or absence of contamination

In this context we define contaminant DNA as sources of DNA that share genetic similarities with the PCR-target, thus might be co-amplified with the sample during the PCR analysis process. While the obvious source is DNA from other sources of the species under consideration, closely related species may also be co-amplified. The extensive, largely unrecorded genetic diversity of environmental microbes, coupled with extensive horizontal gene transfer, makes the recovery of bacterial and fungal sequences especially prone to co-amplification of contaminants.

Why is contamination a problem?

To researchers used to working with the high-quality DNA that can usually be extracted from relatively fresh tissues, it can be

difficult to appreciate the severe effect that contamination can have on samples that contain low template numbers (such as most sources of aDNA). Most laboratories will have small amounts of contaminant DNA in the atmosphere due to the phenomenal amplification efficiency of the PCR reaction. However, in situations where the extracted DNA might contain a large number of template molecules, this contaminant DNA is rarely at high enough levels to cause problems. For example consider an extract containing 20ng DNA, a concentration commonly used for standard PCR. In this case, contamination, which in a well-regulated laboratory will be present at very low levels (for example 10 nuclear DNA copies) will do little to affect the outcome of the result. However, should we now start with an ancient extract that contains only 10 original amplifiable copies, the contamination will be much more serious, representing 50% of the total amplified DNA. Naturally this will have important consequences for the accuracy of the amplified sequence. The situation is even worse for mitochondrial DNA, the molecule of choice for most aDNA studies. This is due to the fact that a single somatic cell contains roughly between 1,000 and 10,000 copies of mitochondrial DNA (mtDNA)[22]. Thus, if an ancient DNA PCR starts from a single copy of mtDNA, not an unusual event[23], then the equivalent of 1/1,000 of the mtDNA content of a single cell is enough to result in a 1:1 ratio between endogenous ancient and modern contaminating DNA.

Methods of spotting contamination

Although a serious problem to ancient DNA studies, there are a number of techniques that can be employed to help gauge whether a sample is contaminated. One method is through the identification of mosaic haplotypes – in essence haplotypes which past experience/information can determine as having arisen as a result of the combination of several phylogenetically unrelated markers[24]. For example, the mtDNA information given on one specimen (Paglicci-12) investigated by Caramelli is not compatible with any known branch in the mtDNA phylogeny (without invoking

recurrent mutations at positions that do not seem to have a high mutation rate) contrary to the authors' claim (the roles of C and T at nucleotide position 10873 in the mtDNA phylogeny were confused)[25]. However, although useful, naturally this method relies on *a priori* knowledge or familiarity with the target, which is not always possible. In addition, it is plausible that mosaic haplotypes might form through severe *post mortem* damage or phantom mutations that had affected the results and might also had found their way into the consensus sequences[26].

An alternative method is through the molecular cloning and subsequent sequencing of amplified PCR products (a method for examining how many different sources of DNA are present within an aDNA extract). When PCR-clones are generated in sufficient number, it can quickly become apparent that the specimen contains numerous distinct sequences. However, difficulties do exist with this method, for example non-consensus mutations (often termed miscoding lesions) might also arise through severe *post mortem* damage or as phantom mutations. In this context it should also be kept in mind that the majority sequence is not necessarily the correct one. As can be seen from the first Neanderthal mtDNA study, depending on the amplification, Neanderthal sequences were both in the majority compared to modern human contamination and so rare that they could not even be detected using primers that amplify both human and Neanderthal mtDNA[27].

Thus, no clear-cut methodology for identifying contamination exists, and especially for studies on ancient human DNA researchers, reviewers and readers need to use common scientific sense to judge each study, and in fact each sequence, on an individual basis.

Although all contamination affects aDNA studies in a similar matter, two clear sources can be distinguished – contaminants can be derived from within the ancient DNA laboratory (external), or be present in the sample prior to the analysis (innate). As there has been a tendency in the aDNA field to regard the second as easily avoidable or treatable, the focus of previous criteria has generally

been on the first. However, it is fundamental to clearly understand the sources and persistence of both forms in order to properly appreciate the contamination-related problems facing the field.

External contamination

External sources of contamination are usually those that arise in the sample/extract as a result of careless laboratory practice. In particular this problem arises as the low levels of DNA surviving in an ancient sample or extract can easily be outnumbered by external sources of good quality/high concentration DNA – such as PCR products, skin cells imparted on the sample during handling or falling in extract, and modern positive controls. However, despite these problems, such contamination can be addressed if suitable equipment and care is used to prevent contact between these sources of DNA and the ancient samples.

Most of the previously published lists include:

- The physical isolation of the DNA extraction and post-amplification areas and one-way transfer of materials from the extraction/PCR set-up area to post amplification areas (to prevent contamination of the extractions with previously amplified DNA)
- The extraction of aDNA in environments not used to extract modern DNA and not using PCR positive controls (to prevent sample cross-contamination).
- The use of appropriate extraction and PCR negative controls (to monitor for the above).
- The use of suitable protective clothing and sterilization techniques (to keep other sources of DNA out of the extract).

Whilst all of these are useful precautions for all PCR work, they do not serve to validate the data, and neither does their absence invalidate the data. The risk presented by external sources of contamination can be estimated in each case by assessing how much of the target organisms DNA is likely to be in the extraction environment, remembering that reagents and consumables used during DNA extractions and PCR-amplification may contain sources of contaminant DNA. Bovine serum albumin, for example,

is often used in PCR reactions to counter PCR inhibition but is also a source of bovine DNA which can affect studies on bovids[28]. Bovine, and to a much larger extent, human DNA is apparently also found in commonly used reaction tubes[29].

Innate contamination

Innate sample contamination, present within the sample prior to laboratory analysis, is a major problem for aDNA studies. Innate contamination arises through sample contact with other sources of DNA: examples include human cells, environmental bacteria and traditional glues[30]. Indirect sources may include animal specimens, pets or traces of food items, cellular material of which is introduced to the sample by handling. The extent of the problem is dependent on three factors - the degree to which the sample will uptake and retain contaminants, the sample's handling and contaminant-exposure history, and the degree to which any contaminant can be differentiated from the sample.

Whether a sample will uptake contaminants or not is a direct effect of its porosity. While DNA sources such as hair appear to be fairly resistant to the uptake of contaminants[31], more porous tissues such as bones and teeth are at particular risk. Fresh bone and dentine is approximately 8% highly interconnected airspace by volume, and in degraded bones this figure can rise to over 40%[32]. While the enamel on a tooth's crown is impermeable to liquid, the dentine-composed root has no such protection, and is frequently in contact with sources of contamination. Human bone seems especially predisposed to contamination in this context; due to the extent of secondary remodeling undergone by human bones (associated with our relatively long lifespan), human bone is more porous than almost all other mammal bone and is far less likely to have been butchered than animal bone, leading to an increase in microbial tunneling[33].

Most methods that are currently used to decontaminate ancient bones and teeth from contaminant DNA only treat the surface of a sample, not accounting for the depth into the sample any

contaminant DNA may be drawn, or the subsequent stabilizing effects of the hydroxyapatite matrix[34]. This includes the exposure of the sample to l=254nm UV light, washing in bleach (12% sodium hypochlorite), physical removal of the surface by sanding or shotblasting and brushing the surface with implements such as toothbrushes.

To summarise, two problems are presented by innate contamination. Firstly, there is a general (erroneous) belief that current aDNA sample pre-preparation techniques successfully decontaminate the samples. Secondly, almost all of the currently adopted authenticity guidelines are geared towards the prevention of external contamination. As a result, if a sample contains sufficiently low levels (or even no remaining) endogenous DNA, and sufficiently high levels of a single, contaminant DNA source, the guidelines may erroneously convey a seal of authenticity on a flawed study.

3. The accuracy of the amplified DNA

The generation of accurate DNA sequences is one of the principal challenges facing aDNA studies. Naturally, whenever several fragments of mtDNA are generated, there is a risk of sample mix-up, incurred by careless handling of sample tubes. Since this kind of artifact appears to be frequent with modern DNA[35], artificial recombination of this type would also be expected to affect aDNA studies in some cases. In addition however, the effects of contamination, *post mortem* miscoding lesions and so-called phantom mutations add further possible problems.

Contamination can lead to the generation of erroneous DNA sequences through the co-sequencing of the authentic and contaminant DNA. Depending on the relative concentrations of each, one of the following might be observed in directly generated DNA sequence (as opposed to that generated through molecular cloning):

a) An aDNA extract that contains one source of contaminant DNA that grossly outnumbers the authentic DNA (and any other

less common contaminants) will generate a sequence that matches the contaminant

b) An aDNA extract that contains similar amounts of the contaminant and authentic DNA, or that contains very low levels of authentic DNA, and equally high levels of several distinct contaminants, will generate sequences that are a composite of the source-sequences (likely containing heteroplasmic sites at positions that differ between the two sequences). Alternatively, due to phenomena such as jumping PCR[36] or other PCR-based forms of recombination, mosaic sequences may be formed.

Various forms of *post mortem* DNA damage are known to generate miscoding lesions that can alter the aDNA sequence that is generated. As with contamination, whether such modifications are likely to affect the final sequence generated or not is related to the quantity of initial PCR-amplifiable templates per PCR reaction. For example, consider damage-derived miscoding lesions. Assume an extract contains 10,000 copies of template, and that at least 50% of the templates must be damaged at a single position for it to be visualized on a direct sequence. As rates of DNA damage *per nucleotide position* are typically low[37] the chance that the same site will be damaged (or misread by the PCR enzyme) 5000 times is effectively zero. Thus in this case DNA damage is unlikely to affect the sequence. However, if the DNA extract contains just one amplifiable molecule, and if 2% of the sequence is damaged, then 2% of all positions will be determined incorrectly with single amplifications.

In addition to damage, a source of non-authentic mutations exist that may convey an impression of *postmortem* damage - phantom mutations. These are sequence modifications that appear to repeatedly target certain nucleotide positions due to biochemical problems in the sequencing process that is conventionally used to recover the DNA sequence from PCR amplified nucleic acids[38].

One simple (although important) technique for dealing with the above problems is the amplifications of a particular sample/region/extract on several occasions, followed by sequencing

in both directions (if a sample is not cloned) as it is deemed unlikely that the same position will be modified in each independent template. However, we warn that nevertheless, damage may still lead to errors in large data sets.

4. Is there enough data to support the conclusions

The retrieval of accurate DNA sequences is not always enough to ensure that results are scientifically interesting. Naturally aDNA studies are hampered by factors that limit the samples that can be obtained, and thus can only provide very select insights into the past. In particular, these problems arise due to biases in which samples are preserved, which sites are excavated, the often un-representative nature of excavated material – including biases towards sexes, social classes and the relatedness of individuals, and even which specimens aDNA researchers are allowed to sample. Whilst these are common themes in archaeology and palaeontology, and there is an extensive and sophisticated literature dealing with them, they are rarely addressed in aDNA studies.

In addition, conclusions from ancient population genetic studies are often very weak as they attempt to use standard population genetic measures on only a handful of ancient individuals. Most modern methods are not designed to cope with small numbers of samples, and lack the statistical power necessary for accurate conclusions to be drawn. One method that has recently become available to partially resolve the issue in some circumstances is through the use of computational methods that employ heterochronous (as opposed to isochronous) sequences to derive population parameters from numbers of temporally dispersed samples[39]. A further problem with using small datasets, is that contamination and other artifacts may only become visible when many ancient individuals are used in the analysis (for example, through the presence of a recurring sequence motif in multiple samples allocated to quite different branches of the mtDNA phylogeny).

However, there will be occasions when single/few specimens are

adequate to answer a particular question. For example, most Neanderthal aDNA studies published to date have relied on sequences from only one or a few specimens[40], but as these studies investigate a taxonomic question of sufficient time depth to allow differentiation of the key taxa, multiple samples are not required to determine the true relationships.

What to consider when appraising aDNA studies

As the above discussion highlights, two problematic types of aDNA studies exist. One group contains those studies were the underlying chemistry seems to reject the data[41]. Based on current knowledge it is not difficult to quickly identify such studies – for example any claims of DNA recovery from samples where the environment is very hot or very hot and damp (*e.g.* ancient Egypt or tropical conditions), or from samples that are dated to over several hundred thousand years old, and are not found in exceptional preservations conditions (that is very cold and dry, such as unthawed glacial ice or permafrost) immediately seem suspect.

The second group of studies, however, represents a much more problematic situation, and contains those studies that offer results that might be correct, but equally, might be wrong. Essentially, these studies are those where insufficient evidence is presented in the paper to make a decision. Unfortunately, we believe that this includes most human aDNA studies published to date. One recent example that we have highlighted before is the publication of two DNA sequences that were reportedly extracted from ~24 KY old human samples[42]. These sequences (from the first hypervariable segment) were extracted and generated adhering to one of the strictest authenticity criteria published to date[43]. The sequences themselves were void of any discriminative features because they matched the basal Eurasian mtDNA sequence motifs (still present in modern mtDNA) and thus matched what would be expected from European samples of 24 KYA (based on current knowledge of the distribution and evolution of European mtDNA sequences)[44]. Hence, as no information was provided on the sample's handling history, when the problems of

innate contamination are considered it becomes impossible for a reader to decide whether the sequences are authentic or contaminant. As in our original article[45], here we stress that we are not implying that the data generated is *definitely* a result of contamination, but we argue that it is impossible for an external party to decide between the two options. Therefore which of the possibilities a reader favors cannot be based on scientific reasoning but only on personal preferences. In fact, if these samples were contaminated with DNA from a single living human, sequences that make perfect sense could be produced with all criteria fulfilled. Nevertheless, the result would still be wrong. This is a general problem in ancient human DNA studies in which the sequences of the researchers involved are likely to be similar or identical to the sequences expected from the fossils. Thus, the general problem with results from such studies is that they therefore provide little – if any – new scientific insights - and as the study of aDNA is both expensive and time-consuming (if undertaken properly), it seems prudent that studies should only be undertaken if the data one might obtain will be useful in answering the questions asked.

Criteria that might be considered when assessing an aDNA study

We hope that, if nothing else, the above discussion highlights to the reader that the study of aDNA is sufficiently complex that strict check-lists for authenticity are not suitable (for example, see Table 1).

Tab. 1 - Comments on some of the authenticity criteria commonly used in aDNA studies

Isolation of work areas - in order to separate sample preparation from PCR amplified products.	Such criteria are important in order to help stop and identify laboratory-derived sources of contamination (such as cross-contamination from previously extracted and amplified samples, contaminants in reagents, DNA shed by the researcher).
Negative control extractions and amplifications - to screen for contaminants entering the process at either stage	
Appropriate molecular	The current knowledge of DNA degradation is be no means complete, thus the

behaviour – due to DNA degradation, aDNA studies should be suspicious of success in amplifying large DNA fragments.	important word here is 'suspicious'. Although many aDNA studies report no successful amplification of fragments over a couple of hundred base pairs in length, in optimal conditions (such as very cold, or fairly young sample), larger (several kb) amplifications may well be possible. Therefore we advocate considering carefully whether anything reported suggests whether large fragments may survive.
Reproducibility – multiple PCR and extractions should yield consistent results	While beneficial in some situations, unfortunately 'Reproducibility' may confer a seal of authenticity upon samples that are contaminated prior to the analysis (*e.g.* through handling). For example, if contamination is homogenously distributed through the sample, and little or no endogenous DNA remains, independent extractions will yield identical results – although results which derive from the contaminant and thus are erroneous.
Cloning of products - to assess for damage, contamination and jumping PCR.	This criteria is useful in some situations, though may not always be necessary. Therefore it can be wrong to reject studies that do not adhere to it. Currently, cloning is one of the only reliable methods that can spot the presence of contamination in aDNA extracts (although, for reasons similar to those outlined in Reproducibility, not always). Thus it is vital in studies that investigate easily contaminated specimens (mainly humans and microorganisms). In addition, the presence of aDNA damage can often help indicate that the DNA is old. However, in situations where it is apparent that high levels of endogenous DNA survive and/or contamination is unlikely, cloning may not be necessary.
Independent replication – the generation of consistent results by independent research groups.	Useful in situations where results are suspect, but hopefully this is not a common experience. In addition, Independent Replication is useful in spotting laboratory-based contaminants that may afflict a particular laboratory or lab-specific artifacts incurred by particular sequencing conditions. However, if a study has employed adequate sample preparation guidelines, the sample is from an already well-characterized species/type, and other factors suggest contamination is unlikely, Independent Replication may be unnecessary. Furthermore, if the sample is contaminated prior to the analysis (as above) Independent Replication can also confer an artificial seal of approval. Furthermore, Independent Replication is usually only applied to a subset of the data (due to financial limitations), which naturally limits its intended use.
Biochemical preservation – preservation of other biomolecules that correlate with DNA survival (*e.g.* collagen, amino acid racemisation) should indicate good sample preservation	Various reports have demonstrated that bone preservation can vary both across archaeological sites, and even within a particular bone (Serre *et al.* 2004). Therefore, good preservation should be taken as supporting evidence, but not proof, of authenticity. Bad preservation should cast serious doubt upon the results.
Quantification – by competitive PCR or Real-Time PCR to give an indication of the number of starting DNA templates in PCR reactions.	Quantification is a controversial technique, in particular as in many cases it is difficult to meaningfully interpret results. Readers should be aware that any quantification assay is valid only for a particular primer pair, and in particular, due to degradation of the endogenous DNA, in an aDNA extract there will be different quantities of different length fragments. Thus if a Real-Time PCR assay targets a 90 b.p. fragment, the results will only indicate the quantity of 90 b.p. fragments of the target that are in the extract. Thus inferences must be drawn with care. Secondly, an accurate sequence can be generated from as little as a single operational molecule (in particular when there is little evidence of miscoding lesions or contamination). Thirdly no published figures exist as to how high contamination loads may be, and doubtless this figure will depend on sample preservation, porosity and its handling history. Furthermore, samples that are well-preserved are likely to contain fairly high levels of DNA, thus are

	at risk of being rejected. Alternatively, samples that do not contain endogenous DNA, but contain a low level of contamination may appear authentic. Therefore, if used, we advocated that the results are interpreted in association with other indicators of data authenticity.
Associated remains – Are associated remains equally well preserved, and do they show evidence of contamination?	There is a significant difference in the structure of human and animal bone, thus it is possible that human samples are more susceptible to contamination and harder to clean than animal samples. Further, unusual (and thus often genetically interesting) human samples might be expected to have undergone more handling than animal samples, thus rendering them more likely contaminated.

Correspondence should be addressed to: mtpg@gmail.com

Instead it has recently been advocated that a reviewer/author/critical reader should consider, on a case-by-case basis whether the evidence presented is strong enough to satisfy authenticity given the problems. However, as with the authors of the previous paper, we stress that this approach should *not* be confused with relaxing the conditions required for obtaining useful, and authentic, aDNA sequences. There are some *sine-qua-non* requirements that aDNA will always require, that include replication (although not always independent), background information on the sample preservation condition and sampling methods, and attempting analyses in the correct environment with appropriate care.

We have previously[46] suggested the following questions that might be asked in order to help answer the dilemma:

1. Do the age, environmental history, and preservation of the sample suggest DNA survival?

2. Does the biomolecular/macromolecular preservation of the sample, the molecular target amplified, the innate nature of the sample and the handling history of the sample suggest contamination is a risk?

3. Does the data offer proof that the sequence is authentic, and not a result of damage, jumping PCR, and contamination? Would patterns in the data suggest other artifacts such as phantom mutations? Do the authors offer sufficient proof that the sequences are authentic?

4. Do the results make sense, and is there enough data to make the study useful?

In essence we suggest that readers of aDNA studies should not attempt to gauge studies by asking themselves, "Which criteria did the authors check off the list?", but rather by asking "What information is presented that makes the results/conclusions believable?", or even better, "Is there any reason to not believe this?". Furthermore we ask readers to extend these questions a little further and ask, "Even if I believe the data to be slightly inaccurate (a contaminant here, a base change there), does this alter the final conclusion of the data?". For example, while the first Neanderthal sequences published[47] may contain a few errors, they do not influence the conclusion that Neanderthal mtDNA falls outside of the variability of modern humans, although they may affect the estimate of coalescence ages by a few ten thousand years. Similarly, if a paper presents a phylogeny of brown and cave bears[48] based on a reasonable number of sequences, if 2% of the bases are incorrect due to damage, the results are unlikely to be greatly affected. However, should the same studies have instead attempted to investigate the genetic diversity among brown or cave bears[49], then the additional (and artificial) variation may change the conclusions significantly. We hope that if such a cognitive approach can be adopted, then some of the controversies that dog the field might, after 20 years of fighting, start to recede.

BIBLIOGRAPHY AND NOTES

Acknowledgements

This article is an expansion on 'Assessing Ancient DNA Studies' originally published by Gilbert *et al*. (2005a). EW and MTPG therefore acknowledge M. Hofreiter, I. Barnes and H.-J. Bandelt for their valuable input in the original paper.

1. HIGUCHI R., BOWMAN B., FREIBERGER M., RYDER O.A., WILSON A.C., *DNA sequences from the quagga, and extinct member of the horse family*. Nature 1984; 312:282-284. HADLY E.A., KOHN M.H., LEONARD J.A., WAYNE R.K., *A genetic record of population isolation in pocket gophers during Holocene climatic change*. Proc Natl Acad Sci USA 1998; 95:6893–96. ORLANDO L., BONJEAN D., BOCHERENS H., THENOT A., ARGANT A., OTTE M., HÄNNI C., *Ancient DNA and the population genetics of cave bears (Ursus spelaeus) through space and time*. Mol Biol Evol 2002; 19:1920-1933. SHAPIRO B., SIBTHORPE D., RAMBAUT A., AUSTIN J., WRAGG G.M., BININDA-EMONDS O.R., LEE P.L., COOPER A.. *Flight of the dodo*. Science 2002; 295:1683,
2. LEONARD J.A., WAYNE R.K., COOPER A., *Population genetics of Ice Age brown bears*. Proc Natl Acad Sci USA 2000; 97:1651-1654. BARNES I., MATHEUS P., SHAPIRO B., JENSEN D., COOPER A., *Dynamics of mammal population extinctions in Eastern Beringia during the last glaciation*. Science 2002; 295:2267-2270.
3. PÄÄBO S., *Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification*. Proc Natl Acad Sci USA 1989; 86:1939-1943. LINDAHL T., *Instability and decay of the primary structure of DNA*. Nature 1993, 362:709-715. HANDT O., HÖSS M., KRINGS M., PÄÄBO S., *Ancient DNA: methodological challenges*. Experientia 1994; 50:524-529; COOPER A., POINAR H., *Ancient DNA: Do it right or not at all*. Science 2000; 289:1139. HOFREITER M., SERRE D., POINAR H.N., KUCH M., PÄÄBO S., *Ancient DNA*. Nat Rev Genet 2001; 2:353-358.
4. GILBERT M.T.P., BANDELT H.J., HOFREITER M., BARNES I., *Assessing ancient DNA studies*. Trends Ecol Evol 2005; 20:541-544.
5. See note 4.
6. VASS A.A., *Beyond the grave – understanding human decomposition*. Microbiol Today 2001; 28:109-192.
7. COLLINS M.J., NIELSEN-MARSH C.M., HILLER J., SMITH C.I., ROBERTS J.P., PRIGODICH R.V., WESS T.J., CSAPO J., MILLARD A., TURNER-WALKER G., *The survival of organic matter in bone: a review*. Archaeometry 2002; 44:383-394,.
8. Cfr. op. cit. nota 3, in particular: LINDAHL T. *Instability and decay of the primary structure of DNA*. Nature 1993, 362:709-715.
9. See note 8.
10. BELL L. S., SKINNER M. F. & JONES S. J. *The speed of post mortem change to the human skeleton and its taphonomic significance*. Forensic Sci Int 1996; 82:129-140. KOON H.E.C., NICHOLSON R.A., COLLINS M.J., *A practical approach to the identification of low temperature heated boneusing TEM*. J Archaeol Sci 2003; 13:1393–99.
11. See note 3, in particular: HOFREITER M., *et al*. *Ancient DNA: Extraction,*

characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci USA 1989; 86:1939-1943.

12. See note 11. SMITH C.I., CHAMBERLAIN A.T., RILEY M.S., COOPER A., STRINGER C.B., COLLINS M.J., *Neanderthal DNA: not just old but old and cold?* Nature 2001; 410:772-773. SMITH C.I., CHAMBERLAIN A.T., RILEY M.S., STRINGER C., COLLINS M.J., *The thermal history of human fossils and the likelihood of successful DNA amplification*. J. Hum. Evol. 2003; 45:203–17.
13. WILLERSLEV E., HANSEN A.J., RONN R., BRAND T.B., BARNES I., WIUF C., GILICHINSKY D., MITCHELL D., COOPER A., *Long-term persistence of bacterial DNA*. Curr Biol 2004; 14:R9–10.
14. COLSON I.B., RICHARDS M.B., BAILEY J.F., SYKES B.C., HEDGES R.E.M., *DNA analysis of seven human skeletons excavated from the Terp of Wijnaldum*. J Archeol Sci 1997; 24:911-917. BURGER J., HUMMEL S., HERRMANN B., HENKE W., *DNA preservation: A microsatellite-DNA study on ancient skeletal remains*. Electophoresis 1999; 20:1722-1728. BARNES I., YOUNG J.P.W., DOBNEY K.M., *DNA-based identification of goose species from two archaeological sites in Lincolnshire*. J Archeol Sci 2000; 27:91-100. NIELSEN-MARSH C, HEDGES R., *Patterns of diagenesis in bone I: Effects of site environments*. J Archeol Sci 2000; 27:1139-1150, GILBERT M.T.P., RUDBECK L., WILLERSLEV E., HANSEN A.J., SMITH C., PENKMAN K.E.H., PRANGENBERG K., NIELSEN-MARSH C.M., JANS M.E., ARTHUR P., LYNNERUP N., TURNER-WALKER G., BIDDLE M., KJØLBYE-BIDDLE B., COLLINS M., *Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy*. J Archaeol Sci 2005; 32:783-795. JANS M.M.E., KARS H., NIELSEN-MARSH C.M., SMITH C.I., NORD A.G., ARTHUR P., EARL N., *In situ preservation of archaeological bone: a histological study within a multidisciplinary approach*. Archaeometry 2002; 44:343-352.
15. See note 8. GEIGL E-M., *On the circumstances surrounding the preservation and analysis of very old DNA*. Archaeometry 2002; 44:337-342. GÖTHERSTRÖM A, COLLINS MJ, ANGERBJÖRN, LIDÉN K. *Bone preservation and DNA amplification*. Archaeometry 2002; 44:395-404.
16. For a overview: PÄÄBO S., POINAR H., SERRE D., JAENICKE-DESPRÉS V., HEBLER J., ROHLAND N., KUCH M., KRAUSE J., VIGILANT L., HOFREITER M., *Genetic analyses from ancient DNA*. Ann Rev Genet 2004; 38:645-679.
17. BADA J.L., WANG X.S., POINAR H.N., PAABO S., POINAR G.O., *Amino acid racemization in amber-entombed insects: Implications for DNA preservation*. Geochim Cosmochim Acta 1994; 58:3131–3135. POINAR H., HÖSS M., BADA J., PÄÄBO S., *Amino acid racemization and the preservation of ancient DNA*. Science 1996; 272:864-866.

18. HÖSS M., JARUGA P., ZASTAWNY T., DIZDAROGLU M., PÄÄBO S., *DNA damage and DNA sequence retrieval from ancient tissue*. Nucleic Acids Res 1996; 24:1304-1307.
19. POINAR H., STANKIEWICZ B., *Protein preservation and DNA retrieval from ancient tissues*. Proc Natl Acad Sci USA 1999; 96:8426-8431.
20. See note 12, in particular: SMITH C.I., *et al*. *Neanderthal DNA: not just old but old and cold?* Nature 2001; 410:772-773. SMITH C.I., CHAMBERLAIN A.T., RILEY M.S., STRINGER C., COLLINS M.J., *The thermal history of human fossils and the likelihood of successful DNA amplification*. J. Hum. Evol. 2003; 45:203–17.
21. WOODWARD S.R., WEYAND N.J., BUNNELL M., *DNA sequence from Cretaceous period bone fragments*. Science 1994; 266:1229-1232. ADCOCK G., DENNIS E., EASTEAL S., HUTTLEY G., JERMELIN L., PEACOCK W., THORNE A., *Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins*. Proc Natl Acad Sci USA 2001; 98:537-542.
22. BOGENHAGEN D., CLAYTON D.A., *The number of mitochondrial deoxyribonucleic acid genomes in mouse l and human HeLa cells*. J Biol Chem 1974; 249:7991-7995.
23. See note 11.
24. BANDELT H.J., SALAS A., LUTZ-BONENGEL S., *Artificial recombination in forensic mtDNA population databases*. Int J Legal Med 2004; 118:267-273.
25. CARAMELLI D., LALUEZA-FOX C., VERNESI C., LARI M., CASOLI A., MALLEGNI F., CHIARELLI B., DUPANLOUP I., BERTRANPETIT J., BARBUJANI G., BERTORELLE G., *Evidence for a genetic discontinuity between Neanderthal and 24,000-year-old anatomically modern Europeans*. Proc Natl Acad Sci USA 2003; 100:6593-6597.
26. MALYARCHUK B.A., ROGOZIN I.B., *On the Etruscan mitochondrial DNA contribution to modern humans*. Am J Hum Genet 2004; 75:920-923.
27. KRINGS M., STONE A., SCHMITZ R., KRAINITZKI H., STONEKING M., PÄÄBO S. *Neanderthal DNA sequences and the origin of modern humans*. Cell 1997; 90:19-30.
28. SHAPIRO B., DRUMMOND A.J., RAMBAUT A., WILSON M.C., MATHEUS P., SHER A.V., PYBUS O.G., GILBERT M.T.P., BARNES I., BINLADEN J., WILLERSLEV E., HANSEN A.J., BARYSHNIKOV D.F., BURNS J.A., DAVYDOV S., DRIVER J.C., FROESE D., HARINGTON C.R., KEDDIE G., KOSINTSEV P., KUNTZ M.L., MARTIN L.D., STEPHENSON R.O., STORER J., TEDFORD R., ZIMOV S., COOPER A., *Rise and fall of the Beringian steppe bison*. Science. Nov 2004; 26,306(5701):1561-5.
29. HUMMEL S., *Ancient DNA typing. Methods, Strategies and Applications*. Springer-

Verlag, 2003.

30. Cfr. op. cit. nota 11. GILBERT M.T.P., CUCCUI J., WHITE W., LYNNERUP N., TITBALL R.W., COOPER A., PRENTICE M.B., *Absence of Y. pestis-specific DNA in human teeth from five European excavations of putative plague victims*. Microbiology 2004; 150:341-354. NICHOLSON G.J., TOMIUK J., CZARNETZKI A., BACHMANN L., PUSCH L., *Detection of bone glue treatment as a major source of contamination in ancient DNA analyses*. Am J Phys Anthropol 118:117-120, 2002.
31. GILBERT M.T.P., WILSON A.S., BUNCE M., HANSEN A.J., WILLERSLEV E., SHAPIRO B., HIGHAM T.F.G., RICHARDS M.P., O'CONNELL T.C., TOBIN D.J., JANAWAY R.C., COOPER A., *Ancient mitochondrial DNA from hair*. Curr Biol 2004; 14 R: 463-464.
32. TURNER-WALKER G., NIELSEN-MARSH C.M., SYVERSEN U., KARS H., COLLINS M.J. *Sub-micron spongiform porosity is the major ultra-structural alteration occurring in archaeological bone*. Int J Osteoarch 2002; 12:407-414.
33. CURREY J. *Bones. Structure and Mechanics*. Princeton, Princeton University Press, 2002. HACKETT CJ. *Microscopical focal destruction (tunnels) in exhumed human bone*. Med Sci Law 1981; 21:243-65.
34. GILBERT M.T.P., *An assessment of the use of human samples in ancient DNA analyses*. D.Phil. Thesis. University of Oxford, 2004.
35. BANDELT H-J. *Etruscan artifacts*. Am J Hum Genet 2004; 75:919-920. BANDELT H-J, PARSON W. *Fehlerquellen mitochondrialer DNA-Datensätze und Evaluation der mtDNA-Datenbank, D-Loop-BASE*". Rechtsmedizin, 2004.
36. PÄÄBO S., IRWIN D., WILSON A., *DNA damage promotes jumping between templates during enzymatic amplification*. J Biol Chem 1990; 265:4718-4721.
37. HOFREITER M., JAENICKE V., SERRE D., VON HAESELER A., PÄÄBO S., *DNA sequences from multiple amplifications reveal artefacts induced by cytosine deamination in ancient DNA*. Nucleic Acids Res 2001; 29:4693-4799. GILBERT M.T.P., WILLERSLEV E., HANSEN A.J., RUDBECK L., BARNES I., LYNNERUP N., COOPER A., *Distribution patterns of post mortem damage in human mitochondrial DNA*. Am J Hum Genet 2003; 72:32-47.
38. BANDELT H.J., QUINTANA-MURCI L., SALAS A., MACAULAY V., *The fingerprint of phantom mutations in mitochondrial DNA data*. Am J Hum Genet 2002; 71:1150–1160. HERRNSTADT C., PRESTON G., HOWELL N., *Errors, phantom and otherwise, in human mtDNA sequences*. Am J Hum Genet 2003; 72:1585–1586.
39. DRUMMOND A., RAMBAUT A., Beast: *Bayesian evolutionary analysis sampling trees*. Available at http://evolve.zoo.ox.ac.uk/beast/. 2003.
40. Cfr. op. cit. nota 27. KRINGS M., GEISERT H., SCHMITZ R., KRAINITZKI H., PÄÄBO S., *DNA sequence of the mitochondrial hypervariable region II from the*

Neanderthal type specimen. Proc Natl Acad Sci USA 1999; 96:5581-5585. OVCHINNIKOV I., GÖTHERSTRÖM A., ROMANOVA G., KHARITONOV V., LIDEN K., GOODWIN W., *Molecular analysis of Neanderthal DNA from the northern Caucasus*. Nature 2000; 404:490-493.

41. See note 21, in particular: WOODWARD S.R., *et al*. *DNA sequence from Cretaceous period bone fragments*. Science 1994; 266:1229-1232. CANO R.J., POINAR H.N., PIENIAZEK N.J., ACRA A., POINAR JR G.O., *Amplification and sequencing of DNA from a 120-135-million-year-old weevil*. Nature 1993; 363:536-538.
42. See note 4, 25.
43. See note 3, in particular: COOPER A., POINAR H., *Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification*. Proc Natl Acad Sci USA 1989; 86:1939-1943.
44. RICHARDS M., MACAULAY V., HICKEY E., VEGA E., SYKES B., GUIDA V., RENGO C., SELLITTO D., CRUCIANI F., KIVISILD T., VILLEMS R., THOMAS M., RYCHKOV S., RYCHKOV O., RYCHKOV Y., GÖLGE M., DIMITROV D., HILL E., BRADLEY D., ROMANO V., CALÌ F., VONA G., DEMAINE A., PAPIHA S., TRIANTAPHYLLIDIS C., STEFANESCU G., HATINA J., BELLEDI M., DI RIENZO A., NOVELLETTO A., OPPENHEIM A., NØRBY S., SANTACHIARA–BENERECETTI S., SCOZZARI R., TORRONI A., BANDELT H.J., *Tracing European founder lineages in the Near Eastern mtDNA pool*. Am J Hum Genet 2000; 67:1251–1276.
45. See note 4.
46. See note 4.
47. See note 27.
48. LOREILLE O., ORLANDO L., PATOU-MATHIS M., PHILIPPE M., TABERLET P., HÄNNI C., *Ancient DNA analysis reveals divergence of the cave bear, Ursus spelaeus, and brown bear, Ursus arctos, lineages*. Curr Biol 2001; 11:200-203.
49. Cfr. op. cit. nota 1, in particular: ORLANDO L. *et al*. Mol Biol Evol 2002; 19:1920-1933. Cfr. op. cit. nota 2, in particular: BARNES I., *et al*. Science 2002; 295:2267-2270. HOFREITER M., CAPELLI C., KRINGS M., WAITS L., CONARD N., MUNZEL S., RABEDER G., NAGEL D., PAUNOVIC M., JAMBRESIC G., MEYER S., WEISS G., PÄÄBO S., *Ancient DNA analyses reveal high mitochondrial DNA sequence diversity and parallel morphological evolution of late pleistocene cave bears*. Mol Biol Evol 2002; 19:1244-1250.

Correspondence should be addressed to: mtpg@gmail.com