Assessing ancient DNA studies

M. Thomas P. Gilbert[1], Hans-Jürgen Bandelt[2], Michael Hofreiter[3] and Ian Barnes[4]

[1]Ecology and Evolutionary Biology, The University of Arizona, 1041 E. Lowell St, Tucson, AZ 85721, USA
[2]Department of Mathematics, University of Hamburg, Bundesstr. 55, 20146 Hamburg, Germany
[3]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany
[4]The Center for Genetic Anthropology, Department of Biology, Darwin Building, University College London, Gower Street, London, UK, WC1E 6BT

The study of ancient DNA has the potential to make significant and unique contributions to ecology and evolution. However, the techniques used contain inherent problems, particularly with regards to the generation of authentic and useful data. The solution currently advocated to reduce contamination and artefactual results is to adopt criteria for authentication. Nevertheless, these criteria are not foolproof, and we believe that they have, in practice, replaced the use of thought and prudence when designing and executing ancient DNA studies. We argue here that researchers in this field must take a more cognitive and self-critical approach. Specifically, in place of checking criteria off lists, researchers must explain, in sufficient enough detail to dispel doubt, how the data were obtained, and why they should be believed to be authentic.

Challenges facing ancient DNA studies

The past 20 years have witnessed the publication of many ancient DNA studies that have investigated a broad range of topics. Ancient DNA techniques have enabled the analysis of extinct species and populations [1–4] and have provided a means of directly tracking temporal genetic changes [5,6]. However, whereas many ancient DNA studies are valuable, a sceptic might argue that their sheer number, and the optimistic tone of regularly published books and review articles [7,8], hides from the average reader the challenges and pitfalls with which the field is filled [9]. Some of these, such as problems of provenance, dating and ultimate interpretation, are shared with other disciplines in which the material record of the past is studied (e.g. archaeology, geology, paleo-climatology and paleontology). However, ancient DNA studies are also subject to problems that are peculiar to the field, which manifest themselves as difficulties in generating sufficient authentic DNA sequences to make a study conclusive. These 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, which increases the risk of preferentially amplifying a contaminant sequence. The problems are not trivial and, over the relatively short history of the field, they have lead to the publication of several high profile, but flawed studies (reviewed in [10]). Here, we discuss how measures currently adopted to ensure the generation of authentic and meaningful data are inherently problematic, and suggest what a future solution might be.

The nine key criteria

The field has predominantly attempted to deal with these problems through the publication of guidelines that are designed to ensure the quality of ancient DNA data and conclusions. Starting from a few relatively simple suggestions (e.g. the use of negative controls; analyses of multiple extracts per sample; and the observation of an inverse correlation between amplification efficiency and size of the amplification product, reflecting the degradation and damage in the ancient DNA template [11]), the guidelines have evolved over time (e.g. [11–13]) into a more detailed and extensive list of requirements, resulting in the well-known nine key criteria of Cooper and Poinar [14] (Box 1). In particular, these authors argue that, in the absence of full compliance to all nine criteria, the reliability and authenticity of results are uncertain.

In light of these arguments, therefore, it might seem somewhat mysterious to the reader that only a few recent

> **Box 1. The nine criteria for authenticity***
>
> **(i) Isolation of work areas**: to separate samples and extracted DNA from PCR amplified products.
> **(ii) Negative control extractions and amplifications**: to screen for contaminants entering the process at any stage.
> **(iii) Appropriate molecular behaviour**: owing to DNA degradation, the successful amplification of large DNA fragments in ancient DNA studies should be treated with caution.
> **(iv) Reproducibility**: multiple PCR and extractions should yield consistent results.
> **(v) Cloning of products**: to assess for damage, contamination and jumping PCR.
> **(vi) Independent replication**: the generation of consistent results by independent research groups.
> **(vii) Biochemical preservation**: preservation of other biomolecules that correlate with DNA survival (e.g. collagen or amino-acid racemization) should indicate good sample preservation.
> **(viii) Quantification**: by competitive PCR or Real-Time PCR to give an indication of the number of starting templates in the reaction.
> **(ix) Associated remains**: are associated remains equally well preserved, and do they show evidence of contamination?
>
> * After [14].

Corresponding author: Gilbert, M.T.P. (mtpgilbert@spymac.com).
Available online 21 July 2005

studies have been published that appear to have adopted all nine criteria [15,16]. Furthermore, not only do most studies not apply all nine, but they also appear to pick and choose criteria at random, giving no explanation as to why some are adopted whereas others are not. This implies that, if, as argued by those who published the criteria, all nine are required to ensure authenticity, the unexplained omission of any single one must render the data and conclusions of such studies (which includes most studies published to date) unreliable.

Such an implication conflicts with the opinions of those who do not follow the guidelines; how might such behavior be reconciled? One proposal is that the problems facing the field are not as serious as has been suggested and, as such, the guidelines are not necessary. However, as any committed ancient DNA researcher can attest, this is not the case, and there are numerous published examples: for instance, the long-term survival and sporadic appearance of contaminant DNA sequences [17]; the difficulty in decontaminating specimens [18–21]; contaminated reagents [22]; sequence modifications at important discriminatory nucleotide positions [23–25]; and the limitations placed on studies owing to DNA degradation [12]. Why, therefore, are the criteria selectively applied, and what is the average reader to make of this contradiction in ancient DNA research?

Authenticity and ancient DNA

The answer to this dilemma stems from one important principle: the criteria were intended to assist in determining the authenticity of a study, but they cannot replace a crucial consideration of the problem. In particular, although the criteria were not meant to be used as a finite checklist that would guarantee authenticity, their use has, in practice, developed into exactly this. The authenticity and reliability of ancient DNA data arise from a complex interplay of several poorly understood areas of knowledge, principally those of DNA damage and contamination, and, as such, no clear-cut answer exists as to what makes a study reliable. That DNA damage and contamination are problematic has been dealt with on several previous occasions (e.g. [10,18,26]) and a sound appreciation of these phenomena is vital to assess an ancient DNA study accurately.

In light of these problems, it becomes apparent why many researchers view the complete application of the criteria as unreasonable and, thus, selectively adopt criteria to fit their situation. Specifically, some might realize that distinct ancient DNA studies face different problems, whereas others decide that the adoption of all criteria might be a waste of time, effort and money. For example, it is widely accepted within the field that contamination affects samples differently. The amount of related DNA in the environment affects the likelihood as to whether samples are contaminated in a way that influences the retrieval of reliable sequences. As such, samples at greatest risk are those from humans, followed by studies investigating paleopathogens and microorganisms for which the modern counterparts might be ubiquitous. By contrast, analyses of well-characterized extinct species (e.g. cave bears and moas) are at much less

of a risk from environmental contamination if PCR-product back-contamination can be controlled.

How modern is ancient DNA from modern humans?

Strict adoption to full criteria lists does not guarantee authenticity and can bring a false appearance of authenticity to problematic results. For example: during 2003, Caramelli *et al.* published two mitochondrial DNA *HVSI* sequences that had been extracted and sequenced from ∼24 000 year-old human samples [16], a study that adhered closely to all nine criteria (at least, as far as the *HVSI* information was concerned). Anticipating potential criticism, Caramelli *et al.* also described exactly how their work complied with the guidelines and, unusually for this area, gave full disclosure of all sequence data. Regardless, a post-publication debate ensued regarding the reliability of the results [27,28]. Essentially, although the obtained sequences agreed with what would be expected from European samples <24 000 years old [29], they also matched the most probable source of contaminants, that is, some of the most common current sequences from Europe.

Herein lies the problem of using strict adherence to criteria as the basis for proof of authenticity: if the samples were contaminated before the analysis, the criteria provide no way to spot that. For example, independent replication of a contaminated sample will yield the contaminant in both replicas. In the above example [16], bearing in mind the contamination problems that relate to studies of old human bone and teeth (including the apparent ease by which they become contaminated [18–21]), this scenario is realistic. In this case, where the expectation was the recovery of sequences that happened to be identical to the most obvious contaminant [16], there seems no clear way of rejecting the possibility of contamination.

One method for the detection of contaminated data is by referring to previously determined biological or other information. An example of this approach is through the identification of mosaic haplotypes; in essence haplotypes which previous knowledge suggests have arisen as a result of the artificial combination of several phylogenetically unrelated markers [30]. In the study by Caramelli *et al.* [16], the combined mtDNA information given for one specimen (Paglicci-12) is incompatible with any known branch of the mtDNA phylogeny (the roles of C and T at nucleotide position 10873 in the mtDNA phylogeny were confused in [16]) and seems to be of a mosaic nature.

A cognitive approach to ancient DNA

What is the solution? It is our opinion that ancient DNA researchers should take a more cognitive approach with regards to assessing the reliability and conclusions of their data. Suggested criteria remain important, and should not be lightly discarded, but we advocate that, in place of planning or assessing studies by using criteria as checklists, consideration should be given on a case-by-case basis as to whether the evidence presented is strong enough to satisfy authenticity given the problems. This places a responsibility on authors to self-assess their work in light of the problems of the field, potentially through the

---

Box 2. Categorizing risk

A potentially useful method for assessing ancient DNA studies is for authors to classify explicitly (with justification) their target samples and sequences into one of four risk categories, which reflect how easy it is to generate erroneous data through contamination. Within the field of ancient DNA studies, it is generally accepted that those of human ancient DNA suffer the most problems (Highest Risk), followed by studies investigating paleopathogens and microorganisms (High Risk), studies of domestic animals and plants (Medium Risk; although some domestic animals that are likely to be easily contaminated, such as cow, pig or chicken, might warrant a higher risk) and, finally, studies of other organisms (Low Risk).

However, although in principle every targeted DNA stretch in an organism would typically belong to one of these risk categories, the system requires flexibility to reflect the nature of ancient DNA. In particular, certain circumstances, which we advocate should be expressed and argued explicitly as part of any ancient DNA study, could lead to a change in category. For instance, although the mtDNA hypervariable regions of Neanderthal specimens are significantly different from modern humans and, thus, might warrant the placement of such studies in a Medium Risk category, other fragments from the Neanderthal mtDNA coding region to be analysed are likely to be so similar to modern DNA that modern contamination would routinely be co-amplified. This would thus place this exercise into the Highest Risk category.

---

categorization of their samples into risk categories (Box 2) and, having done so, to aid readers and reviewers by outlining potential problems, and presenting sufficient information in their publications so as to persuade the critical reader that the data are authentic.

In this way, rather than stating that particular criteria were followed, authors might first describe how the sequences were generated (which might involve minimal description in low-risk categories), and then would discuss the potential problems that are specific to their study, presenting their conclusions based on the data in the context of these problems. For example, in light of the potential contamination issues that face the study by Caramelli *et al.* [16], we suggest that the provision of more information on the handling history of the samples (e.g. how and when the specimens were excavated; how long after excavation samples were taken for DNA analyses; and the handling history of the sample in the intervening period) would have helped the reader to gauge the likelihood of sample contamination and, hence, the reliability of results. Should such data be unobtainable, then we would argue that the scientific usefulness of such studies comes into question. Furthermore, we would argue that future studies should only be undertaken if the data that might be obtained would be useful in answering the questions asked (Box 3).

Questioning ancient DNA studies

To aid the assessment of ancient DNA studies, we suggest that various questions should be asked and answered in light of what is known about the problems facing ancient DNA studies (Box 4). Furthermore, we encourage the extension of these questions a little further so as to ask: 'Even if I believe that the data might be slightly inaccurate, does this alter the final conclusion of the research?' For example, although the first published Neanderthal sequences [31,32] might contain a few

---

Box 3. What do ancient DNA conclusions really tell us?

Researchers of ancient DNA often overlook the fact that the retrieval of genetic information from ancient remains does not guarantee that that information will be useful, or interpreted correctly. The increasing trend for analyses of clusters of human remains in so-called ancient DNA 'population' studies (e.g. [36–38]) is such an example. The diachronic interpretation of ancient samples, as representatives of past ethnic entities that can be directly compared with modern samples of present ethnic or national groups, is done largely in an atheoretical vacuum, similar to the situation quondam in osteology where paleodemography and paleopathology tacitly presupposed that direct relationships existed between statistics calculated from archaeological skeletal series and the health status of past populations ('the osteological paradox' [39,40]). The meagre variation seen in small-sized burial samples seduces the researcher to fill the interpretive gap between reconstructing a complex series of historical processes and measuring genetic differences with stories that seem to allow the data to 'speak for themselves' [41]. Technical shortcomings in analyzing the data would then further exacerbate the problems of inferring spatiogenetic changes in regional 'populations' [42].

---

damage-induced errors, these would not influence the conclusion that Neanderthal mtDNA is likely to fall outside the clade of modern human mtDNA. By contrast, however, such errors might affect the estimates of Neanderthal coalescence ages by a few ten thousand years and, thus, studies that intend to use the data to this end should take this uncertainty into account.

Conclusions

The history of ancient DNA is not as tarnished as many have claimed and there are published studies (particularly those of extinct organisms such as the dodo [33], giant eagle [34] and moa [35]) that did not follow all suggested criteria, but the results of which are likely to be correct. However, there are also many unexpected unpublished results, especially with regards to the study of DNA from human remains. Indeed, one could ask what might be the outcome of an (impending) attempt to analyse the DNA of *Homo floriensis*: if these hominids were modern humans, how would the fragments of their putative mtDNA (if there was any) tell us if the sequence retrieved is authentic? Or, if they were direct *Homo erectus* descendants but modern contaminant mtDNA was amplified, would the DNA then take the lead in interpreting those bones as modern?

We suggest that the solution to these dilemmas will not result from asking 'Which criteria can I check off the list?'.

Box 4. Key questions to ask about ancient DNA

• Do the age, environmental history and preservation of the sample suggest DNA survival?
• Does the biomolecular and/or macromolecular preservation of the sample, the molecular target amplified, the innate nature of the sample and its handling history suggest that contamination is a risk?
• Do the data suggest that the sequence is authentic, rather than the result of damage, jumping PCR, and contamination? Would patterns in the data indicate other artefacts, such as phantom mutations [43]? Do the authors offer sufficient additional proof that the sequences are authentic?
• Do the results make sense, and are there enough data to make the study useful and/or support the conclusions?

Instead, we advocate that readers, reviewers and authors ask 'What information is presented here that makes the results and/or conclusions believable?' and 'Is there any reason to not believe this?'. In short, perhaps a tenth commandment should be added to the nine key criteria: 'Thou shalt interpret the veracity of the data by a critical consideration of all available information'.

References

1 Higuchi, R. *et al*. (1984) DNA sequences from the quagga, and extinct member of the horse family. *Nature* 312, 282–284
2 Hadly, E.A. *et al*. (1998) A genetic record of population isolation in pocket gophers during Holocene climatic change. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6893–6896
3 Orlando, L. *et al*. (2002) Ancient DNA and the population genetics of cave bears (*Ursus spelaeus*) through space and time. *Mol. Biol. Evol.* 19, 1920–1933
4 Shapiro, B. *et al*. (2004) Rise and fall of the Beringian steppe bison. *Science* 306, 1561–1565
5 Leonard, J.A. *et al*. (2000) Population genetics of Ice Age brown bears. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1651–1654
6 Barnes, I. *et al*. (2002) Dynamics of mammal population extinctions in Eastern Beringia during the last glaciation. *Science* 295, 2267–2270
7 Hummel, S. (2003) *Ancient DNA Typing. Methods, Strategies and Applications*, Springer-Verlag
8 Cipollaro, M. *et al*. (2005) Ancient DNA as a multidisciplinary experience. *J. Cell. Physiol.* 202, 315–322
9 Brown, T.A. and Brown, K.A. (2003) Book Review of 'Susanne Hummel (editor) Ancient DNA typing: methods, strategies and applications. *Hum. Genet.* 114, 124–125
10 Pääbo, S. *et al*. (2004) Genetic analyses from ancient DNA. *Annu. Rev. Genet.* 38, 645–679
11 Pääbo, S. (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl. Acad. Sci. U. S. A.* 86, 1939–1943
12 Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* 362, 709–715
13 Handt, O. *et al*. (1994) Ancient DNA: methodological challenges. *Experientia* 50, 524–529
14 Cooper, A. and Poinar, H. (2000) Ancient DNA: do it right or not at all. *Science* 289, 1139
15 Cooper, A. *et al*. (2001) Complete mitochondrial genome of two extinct moas clarify ratite evolution. *Nature* 409, 704–707
16 Caramelli, D. *et al*. (2003) Evidence for a genetic discontinuity between Neanderthal and 24,000-year-old anatomically modern Europeans. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6593–6597
17 Kolman, C.J. (1999) Molecular anthropology – progress and perspectives on ancient DNA technology. In *Genomic Diversity: Applications in Human Population Genetics* (Papiha, S. *et al*., eds), pp. 183–200, Academic/Plenum Publishers
18 Hofreiter, M. *et al*. (2001) Ancient DNA. *Nat. Rev. Genet.* 2, 353–358
19 Richards, M. *et al*. (1995) Authenticating DNA extracted from ancient skeletal remains. *J. Archaeol. Sci.* 22, 291–299
20 Kolman, C.J. and Tuross, N. (2000) Ancient DNA analysis of human populations. *Am. J. Phys. Anthropol.* 111, 5–23
21 Gilbert, M.T.P. *et al*. (2005) Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *J. Archaeol. Sci.* 32, 785–793
22 Eshleman, J. and Smith, D.G. (2001) Use of DNAase to eliminate contamination in ancient DNA analysis. *Electrophoresis* 22, 4316–4319
23 Gilbert, M.T.P. *et al*. (2003) Distribution patterns of *post mortem* damage in human mitochondrial DNA. *Am. J. Hum. Genet.* 72, 32–47
24 Banerjee, M. and Brown, T. (2004) Non-random DNA damage resulting from heat treatment: implications for sequence analysis of ancient DNA. *J. Archaeol. Sci.* 31, 59–63
25 Threadgold, J. and Brown, T.A. (2003) Degradation of DNA in artificially charred wheat seeds. *J. Archaeol. Sci.* 30, 1067–1076
26 Willerslev, E. and Cooper, A. (2005) Ancient DNA. *Proc. Biol. Sci. B* 272, 3–16
27 Abbott, A. (2003) Anthropologists cast doubt on human DNA evidence. *Nature* 423, 468
28 Barbujani, G. and Bertorelle, G. (2003) Were Cro-Magnons too like us for DNA to tell? *Nature* 424, 127
29 Richards, M. *et al*. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67, 1251–1276
30 Bandelt, H-J. (2004) Etruscan artifacts. *Am. J. Hum. Genet.* 75, 919–920
31 Krings, M. *et al*. (1997) Neanderthal DNA sequences and the origin of modern humans. *Cell* 90, 19–30
32 Krings, M. *et al*. (1999) DNA sequence of the mitochondrial hypervariable region II from the Neanderthal type specimen. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5581–5585
33 Shapiro, B. *et al*. (2001) Flight of the dodo. *Science* 295, 1683
34 Bunce, M. *et al*. (2005) Ancient DNA provides new insights into the evolutionary history of New Zealand's extinct giant eagle. *PLoS Biol.* 3, 1–4
35 Haddrath, O. and Baker, A.J. (2001) Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proc. Biol. Sci. B* 268, 939–945
36 Wang, L. *et al*. (2000) Genetic structure of a 2,500-year-old human population in China and its spatiotemporal changes. *Mol. Biol. Evol.* 17, 1396–1400
37 Vernesi, C. *et al*. (2004) The Etruscans: a population-genetic study. *Am. J. Hum. Genet.* 74, 694–704
38 Alzualde, A. *et al*. Temporal mitochondrial DNA variation in the Basque Country: influence of post-Neolithic events. *Ann. Hum. Genet.* (in press)
39 Wood, J.W. *et al*. (1992) The osteological paradox: problems of inferring prehistoric health from skeletal samples. *Curr. Anthrop.* 33, 343–370
40 Wright, L.E. and Yoder, C.J. (2003) Recent progress in bioarchaeology: approaches to the osteological paradox. *J. Archaeol. Res.* 11, 43–70
41 Bandelt, H-J. *et al*. (2003) What molecules can't tell us about the spread of languages and the Neolithic. In *Examining the Farming/Language Dispersal Hypothesis* (Bellwood, P. and Renfrew, C., eds), pp. 99–111, McDonald Institute Monographs
42 Yao, Y-G. *et al*. (2003) Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. *Mol. Biol. Evol.* 20, 214–219
43 Brandstätter, A. *et al*. Phantom mutation hotspots in human mitochondrial DNA. *Electrophoresis* (in press)