Journal of Nordic Archaeological Science 17, pp. 11–14 (2010)

Why study ancient DNA damage?

Jonas Binladen* [1] & Eske Willerslev[1]

** Corresponding author (jbinladen@snm.ku.dk)*

[1] Ancient DNA and Evolution Group, Niels Bohr Institute & Biological Institute, University of Copenhagen, Copenhagen DK-2100, Denmark.

We present an introduction to the known types of damage occurring in ancient DNA templates. The effect of this damage on the recovered sequences is discussed, together with the possibilities for repair.

Keywords: ancient DNA, damage

Ancient DNA and damage

The genetic codes that are carried within organic material can be investigated in order to generate information that cannot otherwise be obtained through conventional macroscopic or microscopic analyses. Insights into past genetics can be obtained using DNA recovered from contemporary (i.e. modern) samples, although these rely to a large extent on the application of theoretical models to the data. Thus they naturally suffer from model-related weaknesses, e.g. through the use of oversimplified or unrealistic parameters. By contrast, the recovery and analysis of DNA from older specimens provides a direct window into the past and allows inferences to be drawn from data generated from samples scattered in time and space. At the most basic level, this might involve the taxonomic identification of a particular specimen. In an evolutionary context, such data can be used to investigate the taxonomic relationship between specimens or for the recreation of past population dynamics (Shapiro et al. 2004). In a human archeological context, ancient DNA (aDNA) has recently been used to identify the appearance, migration and decline of human genetic groups (Haak et al. 2005).

One defining characteristic of ancient DNA is that it is in at least a partially degraded state. The precise level of damage and its biochemical nature varies from one sample to another (Pääbo 1989; Hoss et al. 1996; Hansen et al. 2006; Binladen et al. 2006). In some cases (often cold-preserved specimens) the samples and their DNA are in an excellent state of preservation, while in others the DNA is damaged to such an extent that it can no longer be analysed.

By comparison with the total number of ancient DNA studies that are undertaken, the study of ancient DNA damage is a neglected area despite this being a universal phenomenon in ancient DNA research. Two questions are central to aDNA damage research: How does the damage affect the DNA templates, and thus ultimately the conclusions we draw from the sequences obtained from ancient samples, and can we repair the damage in order to regenerate useful templates? We present in this paper a review of the current state of knowledge with regard to the above questions.

Damage that prevents or limits PCR amplification

The effects of the biochemical damage that ancient DNA is susceptible to can be divided into two distinct groups. The first concerns effects that limit, or even prevent, PCR amplification of DNA. These include depurination reactions that result in breaks in one or both of the DNA strands in the double helix (single or double-strand breaks) (Lindahl 1993), and crosslinking reactions (where the DNA molecule is bound to other molecules, e.g. proteins or other DNA molecules) that cause the PCR polymerase enzymes to abandon or halt their processing of the template

molecule. Secondly, damage can result in the generation of sequences that are not 100% accurate. These cases include the modification of single nucleotides through oxidation and deamination reactions, generating miscoding lesions (Hansen et al. 2001), and jumping PCR, a process that creates artificial chimeric sequences (Pääbo et al. 1990).

Obviously both groups may have severe effects on ancient DNA studies. Amplification-limiting damage can in the most extreme cases prevent the generation of any sequence information at all, and in less severe casesit can simply reduce the total amount of DNA that is available for analysis. In either situation, it provides an ideal platform upon which the specimens can become susceptible to contamination by exogenous modern sources of DNA, and if the level of contamination is high enough, this may simply swamp the authentic DNA in the sample during the subsequent genetic analyses. In this case the resulting genetic data (e.g. DNA sequences or restriction enzyme digests) will be derived from the contaminant and will clearly not represent the true original genetic material of the sample. Within the field of ancient DNA, the study of ancient human and microbial genetics is notoriously susceptible to contamination, because modern human and microbial DNA is ubiquitous in the environment, making it exceedingly difficult to discriminate between the two.

Depurination reactions

For most of the history of ancient DNA it was believed that hydrolytic depurination reactions and the subsequent associated strand breaks were the major damage processes that hindered the amplification of ancient DNA (Lindahl 1993). A recent report (Hansen et al. 2006) has nevertheless questioned this belief. The chemical basis for this effect is the hydrolytic removal of purines, leaving the DNA molecule weakened and ultimately susceptible to cleavage through ß-elimination. This belief was based on *in vitro* experiments to examine the behaviour of free DNA in aqueous solutions (Lindahl & Nyberg 1972), and several subsequently published studies that have attempted to model the long-term survival of DNA have based their arguments almost solely on the assumption that depurination is the predominant form of damage. However, there are problems associated with the assumption that DNA as preserved *in situ* acts in a comparable manner to free DNA in aqueous solutions. Firstly, the temperature of the environment in which the ancient DNA has been present will have varied through time, so that the dynamics of the degradation reactions *in vivo* will have been more complex (Smith et al. 2001) than the constant environments employed in *in vitro* studies. Secondly, the environment in which the ancient DNA was actually preserved could have slowed down the rates of damage through potential protective effects (Salamon et al. 2005).

Crosslinking

The presence of damage caused to ancient DNA by crosslinking has been known for a long time (e.g. Pääbo 1989), but it is only recently that attempt have been made to assess exactly how predominant this is in ancient DNA. One recent study of ancient permafrost soil has demonstrated that it is crosslinks rather than strand breaks that are responsible for the majority of DNA damage (Hansen et al. 2006). If this finding is shown to be a general phenomenon in ancient DNA, it will clearly emphasize how application of the results of *in vitro* damage assays to aDNA studies can be misleading, and the outcomes must be treated with caution.

Damage that results in the generation of erroneous sequences

DNA damage events that result in the generation of erroneous sequences are much easier to observe directly than other forms of DNA damage, as they are *de facto* apparent among the common outcomes of most aDNA studies, the genetic sequences themselves. Those that are most commonly remarked upon are miscoding lesions - damage events that result in the modification of the nucleotide components of DNA in such a way that an incorrect sequence is generated when they are subsequently amplified through PCR. Miscoding lesions are most apparent in cloned sequence data, although they can also be observed as heteroplasmic bases in direct sequence data. In cloned data they are observed as sporadic variations at specific positions within the overall cloned sequence dataset. Damage observed in this way has provided the means by which most miscoding lesion studies have generated their data.

There are several problems with this approach, however. Firstly, the assessment of cloned sequence data can only provide an indirect window onto the damage - what is observed represents the final PCR outcome of the initial damage. As such, biases might

be observed in the damage load if the PCR polymerases prefer to amplify certain types of damage or favour damage-free templates (Stiller et al. 2006).

Secondly, due to the dual strand reverse complementary organization of DNA it is impossible to identify the strand from which the damage occurred. This means the observed damage must be grouped together in complementing pairs, e.g. substitutions of thymine for cytosine (C→T) together with substitutions of adenine for guanine (G→A) (Hansen et al. 2001). As a consequence it becomes difficult to explain the underlying chemical background.

Data from *in vitro* experiments on modern DNA have been used to argue that the deamination of cytosine to uracil, which acts as an analogue for thymine, is the driving force for these C/G→T/A changes. A more direct means of detecting uracil in ancient DNA templates would be to use the enzyme uracil/N-glycosylase (UNG) (Hofreiter et al. 2001).

Novel sequencing-by-synthesis methods have recently lifted the study of miscoding lesions to a level of higher resolution. This method allows the orientation of the original strand to be determined, which enables segregation of the previously studied pairs of complementary bases. In this way Gilbert et al. (2006) and Stiller et al. (2006) have shown that while C→T deaminations are the dominant miscoding lesions in aDNA, They furthermore reported G→A modifications contribute a significant amount of the damage, however Brotherton et al. (2007). showed this finding to be an artefact of the method used and furthermore showed C→T deaminations to be the all dominating miscoding lesion.

Interspecies contaminating sequences are relatively easy to identify, but sequences with miscoding lesions are more difficult to observe as they are homologues of the expected result sequences, but not identical to them.

Repair

Crosslinks

Poinar et al. (1998) have recommended the use of the chemical compound N-phenacylthiazolium bromide (PTB) in aDNA studies. According to their data, this cleaves some of the crosslinks, increasing the yield of DNA recovered from coprolites. By contrast, however, subsequent application of PTB to additional coprolites (Kemp et al. 2006) or bones (Gilbert, unpublished data) has yielded less promising results. It is not known as yet why this discrepancy exists, although it may be due to differences between samples in the levels of the different types of crosslinks.

Single strand breaks

Single strand breaks within a dual strand DNA molecule might in theory be repaired by submitting the DNA to various enzymatic treatments. This is because, although the information carried by the missing nucleotide is absent, a complement is preserved on the opposite strand. Promising attempts to repair single stranded breaks using a combination of the enzymes DNA polymerase I (*E. coli*) and T4 DNA ligase have been reported (Di Bernardo et al. 2002), but this approach has not yet been fully investigated. Furthermore, Hansen et al. (2006) raised the objection that, as the single strand breaks found in aDNA are most likely a result of depurination (resulting in 3'-termini with aldehyde groups), gap-filling by mans of DNA polymerases, (such as DNA polymerase I) will be inhibited. In order to repair the majority of single strand breaks in aDNA templates pre-treatment with T4 polynucleotide kinase (PNK) and AP endonuclease I is required to ensure that most of the 5'- and 3'- ends have PO_4 and OH groups. The effect of the pre-treatment proposed by Hansen et al. on ancient DNA has not yet been tested.

Miscoding lesions

The major form of miscoding lesion, the deamination of cytosine to uracil or its analogues, can be removed from aDNA through incubation with the enzyme uracil/N-glycosylase (UNG) or its close analogues. This treatment excises uracil residues, resulting in the generation of nicks in the DNA template and preventing the subsequent amplification of the template in question. Although no attempts have yet been made to follow UNG treatment with single strand break repair, as detailed above, this remains a potential avenue.

Conclusions

A large amount of basic research needs to be done in order to characterize the damage processes that affect ancient DNA. Although people have been aware of the problem since the early days of work in this field, we are still not aware which types of damage occur or

at what rates, with the exception of the C→T miscoding lesion, and as such are unable to propose more than a few token solutions to the problem. As such, until this knowledge has been generated, we will not be able to understand fully the range and potential of DNA in ancient samples.

Acknowledgement

We wish to thank Dr. Tom Gilbert for many fruitful discussions, for sharing his ideas with us and for his critical reading of this manuscript.

English language revision by Malcolm Hicks.

References

Binladen J., Wiuf C., Gilbert M. T., Bunce M., Barnett R., Larson G., Greenwood A. D., Haile J., Ho S. Y., Hansen A. J. & Willerslev E. 2006. Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics* 172, pp. 733–41.

Brotherton P., Endicott P., Sanchez J. J., Beaumont M., Barnett R., Austin J. & Cooper A. 2007. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Research* 35, pp. 5717–28.

Di Bernardo G., Del Gaudio S., Cammarota M., Galderisi U., Cascino A. & Cipollaro M. 2002. Enzymatic repair of selected cross-linked homoduplex molecules enhances nuclear gene rescue from Pompeii and Herculaneum remains. *Nucleic Acids Research* 30:e16.

Gilbert, unpublished data. In the author's own possession, Tom Gilbert, Center for Ancient Genetics and Environment (AGE), Natural History Museum of Denmark and Department of Biology, University of Copenhagen, and the Greenland Institute of Natural Resources Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark, tgilbert@snm.ku.dk

Gilbert M. T., Binladen J., Miller W., Wiuf C., Willerslev E., Poinar H., Carlson J. E., Leebens-Mack J. H. & Schuster S. C. 2006. Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. Nucleic Acids Research. Sep 14. Epub ahead of print

Haak W., Forster P., Bramanti B., Matsumura S., Brandt G., Tanzer M., Villems R., Renfrew C., Gronenborn D., Alt K. W. & Burger J. 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310, pp. 1016–8.

Hansen A. J., Mitchell D. L., Wiuf C., Paniker L., Brand T. B., Binladen J., Gilichinsky D. A., Ronn R. & Willerslev E. 2006. Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* 173, pp.1175–9.

Hansen A., Willerslev E., Wiuf C., Mourier T. & Arctander P. 2001 Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular Biology and Evolution* 18, pp. 262–5.

Hofreiter M., Jaenicke V., Serre D., Haeseler Av. A. & Pääbo S. 2001 DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research* 29, pp. 4793–9.

Hoss M., Jaruga P., Zastawny T. H., Dizdaroglu M. & Pääbo S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Research* 24, pp. 1304–7.

Kemp B. M, Monroe C. & Smith D. G. 2006. Repeat silica extraction: a simple technique for the removal of PCR inhibitors from DNA extracts. *Journal of Archaeological Science* 33, pp. 1680–1689.

Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362(6422), pp. 709–15. Review.

Lindahl, T., & B. Nyberg, 1972 Rate of depurination of native deoxyribonucleic acid. *Biochemistry* 11, pp. 3610-3618.

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

Pääbo S., Irwin D. M. & Wilson A. C. 1990. DNA damage promotes jumping between templates during enzymatic amplification. *Journal of Biological Chemistry* 265, pp. 4718–21.

Poinar H. N., Hofreiter M., Spaulding W. G., Martin P. S., Stankiewicz B. A., Bland H., Evershed R. P., Possnert G. & Paabo S. 1998. Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Science* 281, pp. 402–6.

Salamon M., Tuross N., Arensburg B. & Weiner S. 2005. Relatively well preserved DNA is present in the crystal aggregates of fossil bones. *Proceedings of the National Academy of Sciences USA* 102, pp. 13783–8.

Shapiro B., Drummond A. J., Rambaut A., Wilson M. C., Matheus P. E., Sher A. V., Pybus O. G., Gilbert M. T., Barnes I., Binladen J., Willerslev E., Hansen A. J., Baryshnikov G. F., Burns J. A., Davydov S., Driver J. C., Froese D. G., Harington C. R., Keddie G., Kosintsev P., Kunz M. L., Martin L. D., Stephenson R. O., Storer J., Tedford R., Zimov S. & Cooper A. 2004. Rise and fall of the Beringian steppe bison. *Science* 306, pp. 1561–5.

Smith C. I., Chamberlain A. T., Riley M. S., Cooper A., Stringer C. B. & Collins M. J. 2001. Neanderthal DNA. Not just old but old and cold? *Nature* 410, pp. 771–2.

Stiller M., Green R. E., Ronan M., Simons J. F., Du L., He W., Egholm M., Rothberg J. M., Keates S. G., Ovodov N. D., Antipina E. E., Baryshnikov G. F., Kuzmin Y. V., Vasilevski A. A., Wuenschell G. E., Termini J., Hofreiter M., Jaenicke-Despres V. & Paabo S. 2006. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proceedings of the National Academy of Sciences USA* 103, pp. 13578–84.