# ancient DNA: damage, contamination, PCR

*Kushal K Dey*

*January 27, 2016*

## Overview

In the first chapter of this Journal Club, we shall consider three very instrinsic features of ancient DNA (aDNA) analysis.

- DNA damage
- DNA contamination
- PCR amplification and sequencing of aDNA

We dicussed this paper

## DNA damage

DNA gets damaged when an organism is alive but there are repair mechanisms that can fix those damages. However, these repair mechanisms do not function after the organism dies. As a result, few intact copies of aDNA tend to survive in old specimens, and those that remain are often highly fragmented and damaged. Preservation in cold environments sometimes prevents damages. Hydrolytic damage leads to single-strand breaks through direct cleavage or following depurination, fragmenting the DNA. Hydrolysis can also cause $C \to T$. DNA damage can be braodly catregorized into following types.

- Strand breaks: occurs due to heat, chemicals, water and microbe degradation. Results in low quantity of surviving DNA with short fragment lengths.
- Miscoding lesions: mainly due to hydrolysis, $C \to T$, $G \to A(xanthine)$, $C \to U$ etc.
- DNA-DNA crosslinks or DNa-protein crosslinks that prevent amplification.

The $G \to X$ transitions are more prominent at the $3^{'}$ end of the strand and $C \to T$ transitions are more prominent at the $5^{'}$ end of the strand. There are enzymes that can detect which of the $U/T$s have come from $C$ and which from $T$, and it can splice up the DNA at these points. Also the DNA fragments we get are shorter in length. There is an increase occurrence in A and G nucleotides near the strand breaks. These features in DNA accumulate over time and may be used to infer the age of the ancient samples. UDG (uracil DNA glycosylase) to remove uracil and splice the ancient DNA. To solve the problem of short fragments, we amplify the overlapping fragments.

We usually have DNA fragments with fragment length $< 100bp$ and contain damaged bases. Cold, dry, temperature-stable environments such as permafrost regions and caves are among the best sources of well-preserved specimens and have permitted large-scale population studies. The theoretical limit to DNA preservation remains between 100,000 and 1,000,000 years.

## DNA contamination

The most serious obstacle to ancient DNA analysis is DNA contamination. The high sensitivity of PCR allows amplification to proceed from only one or a few starting copies of the target sequence, but also often allows contaminating DNA to be amplified. Also, even if the contamination is low, PCR will preferentailly amplify modern DNA against ancient DNA. Copies of the targeted fragment may contain blocking lesions or simply be in low abundance, so that it enters the exponential phase of the PCR many cycles after the reaction has begun. Contamination may occur in different stages.

- Bones and teeth are porous, and contamination may occur via adherence or uptake of exogenous DNA, often from microorganisms residing in the depositional environment.
- Contamination may also occur from modern humans during collection.
- Contamination may occur in lab through laboratory personnel or other reagents. So, contamination may occur at the extraction and amplification stages of the DNA.
- Previously amplified DNA present in the laboratory environment is a particularly dangerous source of contaminating DNA. Even the tiny amount of DNA that is aerosolized when a tube is opened is likely to contain over a million copies of template, that can suppress the ancient DNA completely.

We must also make sure when we sample the ancient DNA from some specimens, that we do not harm the remains or the specimen or destroy parts of it. Conservation is a key issue here. So the type of aDNA sampling used is not destructive.

## Amplification and Cloning

The invention of the polymerase chain reaction (PCR) revolutionized the field of ancient DNA (aDNA) research. In theory, only a single copy of the targeted DNA region is required for PCR, making it a powerful tool for amplifying aDNA from samples where only a handful of intact copies of the target region may remain. PCR is used when we want to make multiple copies of the DNA.

PCR takes as input a DNA. It denatures/separates the two strands by applying heat ($95°C$ for 15 seconds). Heat usually destroys the enzymes that would replicate these strands once they are separated. The solution is cooled to $72°C$ and there is a DNA polymerase known as Taq polymerase which sustains in the heat and helps replicate the DNA strands successfully. Before applying this, we add a primer to each strand. This is the basic philosophy of PCR. This procedure is then repeatedly carried out over each replicate until we get a huge number of replicate DNA fragments. We get an exponential increase in copy number of DNA.

In normal PCR, at a time only one sample of DNA sequence can be amplified. In multiplex PCR, we can simulateneously amplify many targets in one reaction by using multiple primer pairs. Cross primer template binding can hamper this process, so we need to design the primers efficiently. The sequences we take for amplification together should be of least similarity.

Ancient DNA is often highly degraded, and even exceptionally preserved permafrost specimens may contain only 5% of surviving DNA fragments longer than 300 base pairs. Thus when PCR targets DNA sequences with length $> 300bp$,then it would mainly amplify the modern DNA. To overcome this, a series of overlapping primer sets can be used to obtain a long stretch of continuous DNA sequence in small, stepwise fragments. It is also routine to clone at least some of the amplification products of aDNA experiments, as this can identify potential contaminants or PCR artifacts and allow evaluation of the extent of post-mortem damage.

For DNA cloning, we first cut the part of the DNA we want to study. This cutting of DNA is performed by restriction enzymes. These DNA cuts are attached to carriers called vectors (molecules that can replicate and when they replicate recreate the fragment inserted). Vectors are inserted in host cell/living cell in a bacterium (in vivo). Then the vector will find right condition to replicate and thus creates moreclones of the

DNA fragment we want. To see differences between DNA cloning and DNA amplification by PCR, check this site.

However, each PCR requires at least one template molecule of the desired genomic region. Depending on DNA preservation, more or less DNA extract will be required to begin each amplification reaction. Multiple amplifications of short, overlapping fragments may be necessary to reconstruct long, informative DNA sequences. This process is performed in replicate to account for possible sequence errors due to miscoding lesions in the template molecule.

For many studies, even the smallest lane on any next generation sequencing (NGS) instrument will produce excessive amounts of sequence data for a single sample. If larger numbers of samples are to be analyzed, the cost soon becomes prohibitive if a full single lane is used per sample. It is therefore important to be able to pool multiple samples and sequence them in a single lane. As the information about sample origin of individual sequence reads is lost in all NGS approaches, this requires barcoding techniques, in which a specific tag is attached to all DNA fragments allowing them to be sorted bioinformatically after sequencing.

The most efficient way to produce a barcoded sequencing library is to amplify a genomic target region using polymerase chain reaction (PCR) with target-specific primers that include a sequencing adapter and barcode tail.