Engineered polymerases amplify the potential of ancient DNA

Beth Shapiro

Department of Biology, 326 Mueller Laboratory, The Pennsylvania State University, University Park, PA 16801 USA

Corresponding author: Shapiro, B (beth.shapiro@psu.edu).

The generation of genomic data from mammoths and Neanderthals has reinvigorated discussion about whether extinct species could be brought back within the foreseeable future. However, post-mortem DNA decay rapidly reduces the number and quality of surviving DNA fragments, consequently increasing rates of sequencing error and forming a significant obstacle to accurate sequence reconstruction. Recent work has shown that it is possible to engineer a polymerase capable of using even highly damaged fragments as template sequences.

Introduction: the promise of 'ancient' DNA

Since the first successful amplification of DNA from an 'ancient' specimen 25 years ago [1], 'ancient DNA' (aDNA) has attracted interest from across scientific disciplines and from the public. Genetic data extracted from fossil remains promised to solve questions about relationships between extinct and extant organisms [2–4] and provide a window into time to measure evolution as it occurred [5]. The ensuing burst of publications in the early 1990s, many of which have now been discredited (e.g. [6,7]), even led to perhaps overly enthusiastic claims by some popular media that extinct species could be brought back to life. There is little doubt that the ability to amplify sequences from fossil remains has offered a valuable complement to studies of modern DNA. As the field has matured, however, so has our understanding of its limitations, in particular limitations imposed by the damage that accumulates in genetic material over time. Recent work in Phil Holliger's research group has shown that it might be possible to reverse some of the effects of damage-associated lesions in ancient specimens, which could have significant implications for the temporal and geographical range of samples available for analysis [8].

The effect of damage on DNA sequences

Organismic DNA begins to decay immediately after cell death, first by the action of endogenous nucleases, then by external physical and chemical abrasion, for example by water, UV light and the action of bacteria and fungi in the environment. The specific end-result is situation-dependent due to the complex interaction between decay processes. DNA damage can manifest as single- or double-strand breaks, protein crosslinks, the incorporation of structural modifications, such as hydantoins, and the development of miscoding lesions [9]. The rate at which damage accumulates varies depending on environmental conditions. Optimal circumstances for DNA preservation include constant low temperatures, rapid post-mortem desiccation or high salt concentrations. In any environment, the cumulative effect of these processes over time is that no amplifiable DNA endogenous to the sample survives. Previous experimental work has suggested that at 15°C, recoverable DNA will survive for around 100 000 years [9], whereas experimental observations on deep frozen DNA suggest an upper limit for survival of ≈1 million years [10].

Prior to the release of high-throughput sequencing-by-synthesis platforms, most aDNA studies relied on the amplification of selected sequence fragments via polymerase chain reaction (PCR). The replicative enzymes used in PCR, generally heat-resistant polymerases derived from bacteria in the genus *Thermus* (most commonly *T. aquaticus, or 'Taq'*), are limited in their ability to sequence through damaged sites. As the number and condition of potential template molecules declines, the possibility that incorrect nucleotides are incorporated during PCR increases, which can significantly mislead phylogenetic and genealogic analyses [11]. To minimize this problem, laboratory protocols, such as bacterial cloning and replication (Figure 1), are often used to corroborate the accuracy of consensus sequences [12]. Additionally, analytical tweaks, such as the delta statistic [13], allow some uncertainty around the 'true' sequence in phylogenetic analyses. Although such protocols will limit the effect of sequence error, it is clear that increasing the amount and quality of DNA available for PCR amplification would have a considerable beneficial impact on aDNA research.

Making more from nothing: a polymerase capable of reading damaged DNA

Recently, two European research teams demonstrated that polymerases could be bred that are capable of sequencing through two common forms of damage in aDNA: abasic sites (often manifested as single-strand breaks) and hydantoins generated via oxidation [8,14]. Both studies focused on evolving a polymerase capable of extending DNA from a primer sequence containing mismatched bases on the 3′ end. Such mismatches mimic the distorted DNA structures resulting from damage and are known to cause significant stalling and replication failure with standard *Taq*. Using molecular breeding, d'Abbadie *et al.* [8] combined three A-family polymerases (DNA pol I) from *Thermus* to create a hybrid library for compartmentalized self-replication selection. They identified several novel polymerases that outperformed *Taq* in sequencing damaged DNA and that were capable of sequencing through up to four mismatched bases. In a

similar approach, Gloeckner *et al.* [14] used a previously described library to identify a *Taq* variant capable of extending although base mismatches and further characterized this variant to suggest that a single amino acid substitution (M747K) was responsible for this activity.

Most exciting for the future of aDNA was the result of experiments performed by d'Abbadie *et al.* on 47 000- and 60 000-year-old cave bear specimens. Although *Taq* outperformed the novel hybrid polymerases on modern DNA, the hybrids consistently outperformed *Taq* when amplifying aDNA and were even able to amplify from concentrations where *Taq* failed entirely [8]. This confirmed the long-held suspicion that significant amounts of DNA were surviving in ancient specimens but were rendered inaccessible to conventional polymerases by damage. Although it is unclear how the hybrid polymerases perform in comparison to the specialized polymerases normally used in aDNA (e.g. Applied Biosystems AmpliTaq Gold® [15]; Invitrogen Platinum®Taq HiFi [16]), this discovery is key to the future of aDNA because it creates opportunities for research into methods for accessing this previously untapped resource, including further refinements in polymerase activity.

Room for improvement

One drawback to the enzymes described above is how they replace missing information: although each appears to preferentially insert adenine residues at empty sites, neither study presents evidence of any clear pattern. This might be less preferable to the standard practice of visually inspecting multiple cloned PCR products because DNA damage is known to manifest in particular ways and can therefore be readily identified [11]. It could be argued that if sufficient quantities of template molecules are accessible and damage occurs randomly, the resulting consensus sequence is likely to be correct. However, another common form of aDNA damage is miscoding lesions, often formed by deamination of cytosine to uracil [17], that do not interfere with replication and are often only recognizable as errors after several cycles of replication and cloning. The consensus sequence can therefore not be assumed to reliably provide the accurate sequence at all sites.

One way around this problem might be to use an enzyme that preferentially incorporates the correct nucleotides opposite sites of DNA lesions. In a similar experiment to those described above, McDonald *et al.* [18] identified novel chimeric Y-family polymerases similar to DPo4 found in Archaea. Y-family polymerases are known for their ability to sequence through a wide variety of DNA lesions, and the DPo4-like polymerases were shown to be capable of extending both undamaged DNA and DNA with lesions caused by mismatched or baseless sites and/or depurination-derived hydantoins and to be able to preferentially incorporate the correct nucleotides opposite the damaged bases. Problematically, Y-family polymerases show significantly less sequence fidelity than do A-family polymerases, which presents a serious disadvantage to their widespread use in PCR. To circumvent this, McDonald *et al.* used a combination of Y-family and standard A-family polymerase in PCR of DNA sequences damaged by UV irradiation. Remarkably, when used with either AmpliTaq Gold® or standard *Taq* polymerase, the combination of A- and Y-family polymerases outperformed any A-family polymerase by itself [18]. Although this method seems particularly promising, further work will need to be done to characterize the effect of the lower-fidelity polymerase on the distribution of sequencing errors.

Finally, a recent extensive survey of DNA preserved in ice led to the discovery that interstrand crosslinks formed between reducing sugars and amino acid groups might be more common in ancient samples than are lesions caused by depurination [19]. Like other forms of damage, protein crosslinks block polymerase activity, adding to the portion of surviving DNA fragments that are inaccessible for PCR. Although it has been shown that some crosslinks can be broken down by the chemical *N*-phenacylthiazolium bromide [20] (but see [21]), it remains to be seen how much and what effect this type of damage has on the resulting sequences.

Future challenges for ancient DNA research

Our ability to accommodate and even repair damage in ancient specimens has improved considerably in the last few years. However, it is clear that more work remains to be done before the true potential of aDNA can be realized. Although it has not yet been considered here, I cannot underemphasize the importance of another, related source of error that plagues aDNA research: contamination by modern DNA. When the number of starting template copies is small and when the accessible copies are damaged to any extent, the polymerase is likely to preferentially amplify copies of modern DNA that have been introduced to the ancient sample, either from handling of the sample itself at any stage of the experimental procedure or from previously amplified fragments present in the laboratory environment [22]. This is particularly problematic when the potential contaminating sequence is evolutionarily close to the ancient specimen, as in studies involving ancient humans [23] and Neanderthals [24]. With particular relevance to the increasing number of genome sequencing projects underway, distinguishing between observed nucleotide polymorphisms resulting from DNA damage, modern contamination and true differences between extinct and extant taxa is one of the most pressing challenges in aDNA today. Finding an adequate solution, whether by developing novel experimental techniques or identifying computational solutions, will be fundamental to achieving those early promises of aDNA.

Disclosure statement

The author states that they have no conflict of interest to disclose.

References

1 Higuchi, R. *et al.* (1984) DNA-Sequences from the quagga, an extinct member of the horse family. *Nature* 312, 282–284

2 Shapiro, B. *et al.* (2002) Flight of the dodo. *Science* 295, 1683

3 Noonan, J.P. *et al.* (2006) Sequencing and analysis of Neanderthal genomic DNA. *Science* 314, 1113–1118
4 Green, R.E. *et al.* (2006) Analysis of one million base pairs of Neanderthal DNA. *Nature* 444, 330–336
5 Shapiro, B. *et al.* (2004) Rise and fall of the Beringian steppe bison. *Science* 306, 1561–1565
6 Woodward, S.R. *et al.* (1994) DNA-sequence from Cretaceous period bone fragments. *Science* 266, 1229–1232
7 Hedges, S.B. and Schweitzer, M.H. (1995) Detecting dinosaur DNA. *Science* 268, 1191–1192
8 d'Abbadie, M. *et al.* (2007) Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* 25, 939–943
9 Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* 362, 709–715
10 Willerslev, E. *et al.* (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* 317, 111–114
11 Gilbert, M.T. *et al.* (2003) Distribution patterns of postmortem damage in human mitochondrial DNA. *Am. J. Hum. Genet.* 72, 32–47
12 Gilbert, M.T. *et al.* (2005) Assessing ancient DNA studies. *Trends Ecol. Evol.* 20, 541–544
13 Ho, S.Y. *et al.* (2007) Bayesian estimation of sequence damage in ancient DNA. *Mol. Biol. Evol.* 24, 1416–1422
14 Gloeckner, C. *et al.* (2007) Evolving a thermostable DNA polymerase that amplifies from highly damaged templates. *Angew. Chem. Int. Ed. Engl.* 46, 3115–3117
15 Hofreiter, M. *et al.* (2007) Sudden replacement of cave bear mitochondrial DNA in the late Pleistocene. *Curr. Biol.* 17, R122–R123
16 Barnes, I. *et al.* (2007) Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Curr. Biol.* 17, 1072–1075
17 Hofreiter, M. *et al.* (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* 29, 4793–4799
18 McDonald, J.P. *et al.* (2006) Novel thermostable Y-family polymerases: applications for the PCR amplification of damaged or ancient DNAs. *Nucleic Acids Res.* 34, 1102–1111
19 Hansen, A.J. *et al.* (2006) Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* 173, 1175–1179
20 Poinar, H.N. *et al.* (2001) A molecular analysis of dietary diversity for three archaic Native Americans. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4317–4322
21 Rohland, N. and Hofreiter, M. (2007) Comparison and optimization of ancient DNA extraction. *Biotechniques* 42, 343–352
22 Pääbo, S. *et al.* (2004) Genetic analyses from ancient DNA. *Annu. Rev. Genet.* 38, 645–679
23 Sampietro, M.L. *et al.* (2006) Tracking down human contamination in ancient human teeth. *Mol. Biol. Evol.* 23, 1801–1807
24 Wall, J.D. and Kim, S.K. (2007) Inconsistencies in Neanderthal genomic DNA sequences. *PLoS Genet.* 3, 1862–1866

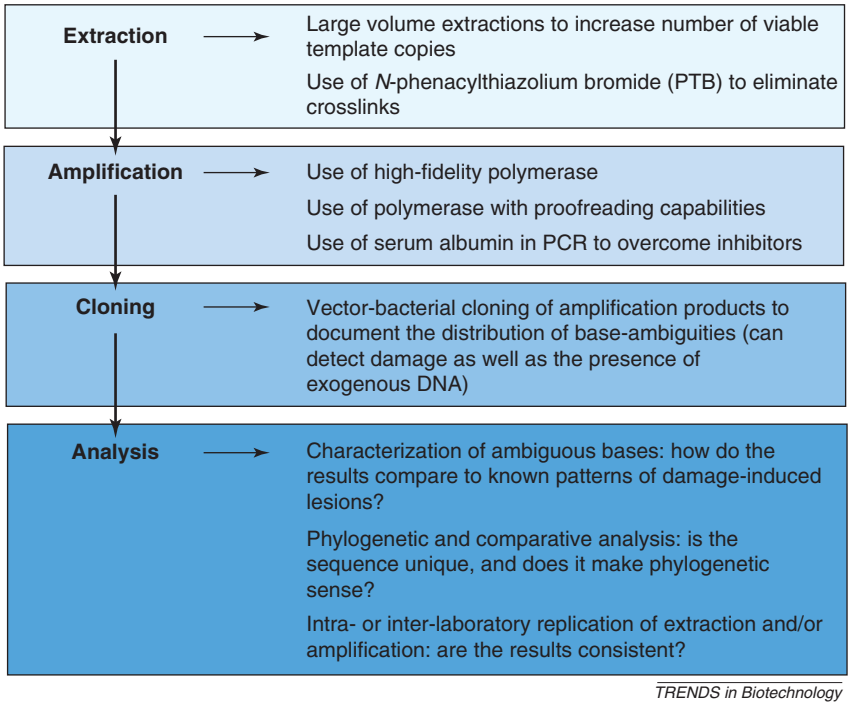

Figure 1. Methods to avoid and detect damage in ancient DNA specimens.