Minireview

Ancient genomes

A Rus Hoelzel

Address: School of Biological and Biomedical Sciences, Durham University, South Road, Durham DH1 3LE, UK. E-mail: a.r.hoelzel@dur.ac.uk

Published: 1 December 2005

Genome ***Biology*** 2005, **6:**239 (doi:10.1186/gb-2005-6-12-239)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2005/6/12/239

Abstract

Ever since its invention, the polymerase chain reaction has been the method of choice for work with ancient DNA. In an application of modern genomic methods to material from the Pleistocene, a recent study has instead undertaken to clone and sequence a portion of the ancient genome of the cave bear.

Nearly 20 years ago a researcher from down the hall in the Department of Genetics at Cambridge University brought in a home-built contraption for us to try. It consisted of a metal box containing a single halogen lamp with a copper sheet on top with eppendorf-tube-shaped indentations, together with a BBC Micro computer to run it. It was a proto-PCR machine, and along with their home-brewed Taq polymerase it worked quite well. This was soon after Saiki *et al.* [1] had described the full method, including the use of a thermostable polymerase, and before long manufacturers were offering a dazzling array of PCR machines (see [2]), including temperature-cycling ovens and water baths (neither of which survived the test of time). A few months after we had incorporated the copper contraption into our methodology, someone from Archaeology brought along a stingless bee, trapped in amber for millions of years. At the time, many people saw that PCR could provide a window into the past, and there was considerable excitement about its potential: even if the number of surviving intact DNA templates was low, they could, in theory, be amplified by PCR. But as with the design of thermo-cycling machines, there was a learning curve. Early results from material more than 1 million years old could not be replicated, and it began to appear that there are limits to how long we can expect intact DNA to survive. While amplifications of material up to hundreds of thousands of years old have been confirmed (for example, from bacteria in permafrost [3]), amplifications of older material remain controversial (reviewed in [4]).

DNA degrades by processes such as oxidation and hydrolysis, leading to lesions that eventually break DNA down into smaller and smaller fragments [5,6]. As this reduces the available template, and given that PCR has the potential to amplify from even a single copy of a genome, contamination with modern DNA template becomes a major problem. One recent study used quantitative PCR to assess the relative proportion of true template to contaminating human DNA in extracts from 5,000-year-old canine bones [7]. Standard precautions were taken, such as the isolation of the work in a separate, ultraviolet-irradiated lab, and cleaning the exterior of the sample before drilling into the bone. Even so, a large proportion of extracts were contaminated with human DNA. Malmström *et al.* [7] found that although the proportion of human template in the extracts varied among the 29 dog bones analyzed, human DNA was always more abundant than canine DNA, in some cases by two orders of magnitude. The template being amplified in this study was the mitochondrial genome (mtDNA). From the earliest ancient amplifications that have stood the test of time through to recent applications this has tended to be the genome of choice. The reason is obvious: there are just two copies of each 'single-copy' nuclear gene in diploid organisms, but there are 2-10 mitochondrial genomes in each mitochondrion, and hundreds or thousands (depending on the cell type) of mitochondria per cell. Given so many copies, the chance of finding intact mtDNA is greater. Contaminating material will, however, also have abundant mtDNA.

Studies of ancient DNA have often involved the amplification of short mtDNA segments for either phylogenetic or, less frequently, population studies. The PCR primers can be designed to amplify specifically from the target species,

thereby helping to reduce the risk of contamination (unless the ancient material is human). A major break with this trend was published by Willerslev *et al.* [8]. The chosen template was from the 18S gene in the ribosomal DNA (rDNA) family. This sequence is highly conserved, and primers were designed to amplify from as broad a range of taxa as possible. The material for amplification was from 2,000- and 4,000-year-old segments of an ice core, from which total DNA was extracted. The amplified DNA was cloned, and the clones sequenced to assess the diversity of organisms found in the two layers of ancient ice. A remarkable diversity of species was revealed in what was a significant departure from earlier applications, combining PCR and cloning strategies.

Now, in the modern era of genomics, there has been a return to basics. The first ancient DNAs to be examined came from ample material that was cloned directly into a vector (no PCR); this material came from a 150-year-old extinct horse called the quagga [9] and from an Egyptian mummy [10]. This was after the concept of PCR had been invented by Kary Mullis in 1983, but before the first patent application by Cetus (where Mullis worked) and the first publication [11]. In a study just published in *Science* by Noonan *et al.* [12], an approach exclusively based on cloning has been used again. This time two metagenomic libraries were constructed by anonymously cloning all DNA present in the samples. One library was from a 44,000-year-old bone and the other from a 42,000-year-old tooth from the cave bear (*Ursus spelaeus*). Creating a genomic library from sub-fossil material is not a very efficient process. In true genomics style, 9,035 clones (1.06 Mb) and 4,992 clones (1.03 Mb) were sequenced from the bone and tooth libraries, respectively. Among these sequences, 1.1% from the tooth library and 5.8% from the bone library were identified as cave bear. The rest were a mixture of unknown (62-66% - the largest portion), local environmental contaminants (11-17%), prokaryotic sequences (17-26%) and other eukaryotic sequences (0.7%).

The strategy is illustrated in Figure 1. DNA was extracted using a silica-based recovery method, and mtDNA copy number estimated by quantitative PCR. Approximately 15 million mtDNA fragments of 100 bp in length were estimated to be in the 25 μl extract. Assuming a ratio of about 1,000:1 mtDNA to nuclear, this suggested roughly 15,000 nuclear copies - in theory enough for a library with 10-fold coverage of the cave bear genome. Extracted DNAs were then end-repaired in preparation for blunt-end ligation into the cloning vector pMCL200, and cells transformed by electroporation. Cloned sequences were screened against existing databases, crucially including the dog genome, which is accessed by the Dog Genome Browser through the University of California at Santa Cruz [13]. Fragments with homology to the dog genome (92% similarity on average) typically comprised just part of an insert. To confirm that these represented cave bear clones, PCR primers were constructed from

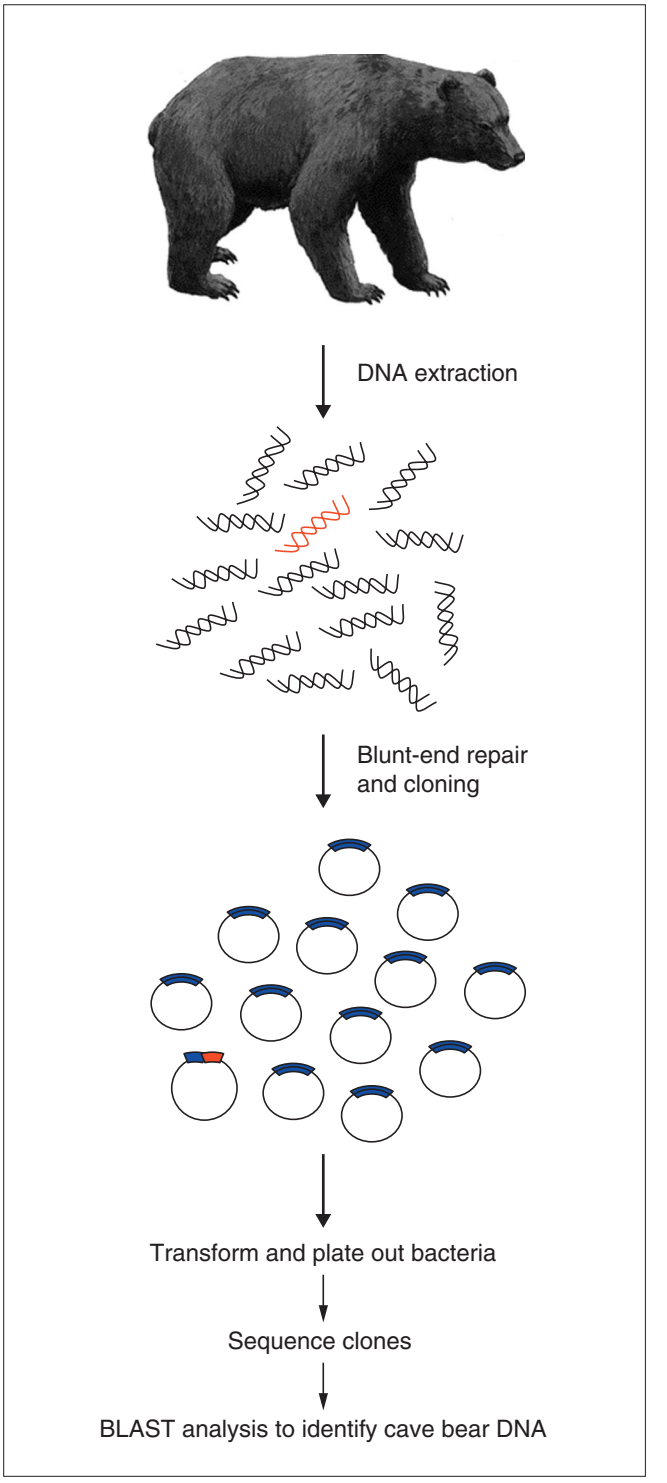

Figure 1
Cloning strategy for constructing and analyzing metagenomic libraries using DNA extracted from cave bear tooth and bone (adapted from [12]). Total DNA from the samples, which will include numerous environmental contaminants, is isolated and cloned. After the cloned DNA was extracted and sequenced, the sequences are analyzed using BLAST against available databases to distinguish the relatively few cave bear clones (red) from the rest of the DNA. In this case, a similarity to canine DNA (unlikely to be a contaminant) was used to identify candidate cave bear DNA.

the insert sequences and used to test for homology with amplified modern brown bear DNA. This matched with identities of at least 97%, suggesting that the clones represented authentic cave bear DNA.

After database screening, 6,775 bp from the tooth library and 20,086 bp from the bone library were attributed to the cave bear. Among these, 4-6% represented reference exon sequences, 4-6% were conserved noncoding, 35-45% were repetitive DNA, and the majority (45-55%) were unannotated sequences. These proportions are roughly comparable to those seen in modern mammalian genomes. There were no matches to mtDNA among the 389 putative cave bear clones, but this is not surprising given the difference in the size of the two genomes (16.5 kb mitochondrial versus 3 Gb nuclear), in spite of the high copy number of the mitochondrial genome.

A phylogenetic reconstruction was then generated using 3,201 bp of the sequences assigned to the cave bear and comparing them with modern black bear, brown bear and polar bear sequences. The estimated substitution rate for the cave bear, based on this phylogenetic tree, was higher than expected. The probable cause was an increased occurrence of GC-AT transitions, given the propensity for deamination to convert cytosine to uracil in ancient DNA. Most of the problem seemed to lie with a few damaged clones in one of the libraries, and when they were removed the substitution rate appeared more consistent with expectation. The same four species (among others) had previously been compared by another group using the mtDNA control region and the cytochrome B (*cytb*) loci [14], and that tree was topologically equivalent to the nuclear DNA tree. This to some extent begs a question: why clone and sequence 2 million base pairs of nuclear DNA when direct amplification of mtDNA (the 'traditional' way) gives you the same result, especially for such a straightforward phylogenetic question? I think the answer is in large part that the construction of this phylogeny was not the only object of the exercise. Instead, these researchers have set their sights on a contemporary of the cave bear, and this project illustrates the potential and scope of the cloning method towards that future objective.

In July of this year, collaborators in the cave bear project from the Max-Planck Institute for Evolutionary Anthropology, and the University of California at Berkeley, announced a joint venture to sequence the genome of *Homo neanderthalensis*. Such a project will raise special problems of its own. The closeness of the human and Neanderthal genomes will make distinguishing Neanderthal from contaminating human clones that much harder. Krings *et al.* [15] showed that human and Neanderthal share 92.5% of their mtDNA control region sequence. Furthermore, some studies have demonstrated the difficulty of identifying specific mutations in ancient samples, given the propensity for mutagenesis in ancient DNA [16]. If successful, however, the Neanderthal genome project should teach us a lot about what it is to be human, and together with the results of the human genome project, something about how we got there.

References

1. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: **Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.** *Science* 1988, **239:**487-491.
2. Hoelzel AR: **The trouble with PCR machines.** *Trends Genet* 1990, **6:**237-238.
3. Kim S, Soltis DE, Soltis PS, Sue Y: **DNA sequences from Miocene fossils: an *ndhF* sequence of *Magnolia latahensis* (Magnoliaceae) and an *rdcL* sequence of *Persea psuedocarolinensis* (Lauraceae).** *Am J Bot* 2004, **91:**615-620.
4. Willerslev E, Cooper A: **Ancient DNA.** *Proc R Soc Lond B* 2005, **272:**3-16.
5. Gilbert MT, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, Cooper A: **Characterization of genetic miscoding lesions caused by postmortem damage.** *Am J Hum Genet* 2003, **72:**48-61.
6. Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S: **DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA.** *Nucleic Acids Res* 2001, **29:**4793-4799.
7. Malmström H, Storå J, Dalén L, Holmlund G, Götherström A: **Extensive human DNA contamination in extracts from ancient dog bones and teeth.** *Mol Biol Evol* 2005, **22:**2040-2047.
8. Willerslev E, Hansen AJ, Christensen B, Steffensen JP, Arctander P: **Diversity of Holocene life forms in fossil glacier ice.** *Proc Natl Acad Sci USA* 1999, **96:**8017-8021.
9. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC: **DNA sequences from the Quagga, an extinct member of the horse family.** *Nature* 1984, **312:**282-284.
10. Pääbo S: **Molecular cloning of ancient Egyptian mummy DNA.** *Nature* 1985, **314:**644-645.
11. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N: **Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle-cell anemia.** *Science* 1985, **230:**1350-1354.
12. Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Detter JC, Pääbo S, Rubin EM: **Genomic sequencing of Pleistocene cave bears.** *Science* 2005, **309:**597-600.
13. **Dog (*Canis familiaris*) Genome Browser Gateway** [http://genome.ucsc.edu/cgi-bin/hgGateway?org=Dog&db=0&hgsid=34526766]
14. Loreille O, Orlando L, Patou-Mathis M, Philippe M, Taberlet P, Hanni C: **Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages.** *Curr Biol* 2001, **11:**200-203.
15. Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S: **Neanderthal DNA sequences and the origin of modern humans.** *Cell* 1997, **90:**19-30.
16. Pusch CM, Broghammer M, Nicholson GJ, Nerlich AG, Zink A, Kennerknecht I, Bachmann L, Blin N: **PCR-induced sequence alterations hamper the typing of prehistoric bone samples for diagnostic achondroplasia mutations.** *Mol Biol Evol* 2004, **21:**2005-2011.