Beringia as an Ice Age genetic museum
Beth Shapiro and Alan Cooper*

Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, Oxford University, South Parks Road, Oxford, OX1 3PS UK

Received 4 November 2002

Abstract
Thousands of Late Pleistocene remains are found in sites throughout Beringia. These specimens comprise an Ice Age genetic museum, and the DNA contained within them provide a means to observe evolutionary processes within populations over geologically significant time scales. Phylogenetic analyses can identify the taxonomic positions of extinct species and provide estimates of speciation dates. Geographic and temporal divisions apparent in the genetic data can be related to ecological change, human impacts, and possible landscape mosaics in Beringia. The application of ancient DNA techniques to traditional paleontological studies provides a new perspective to long-standing questions regarding the paleoenvironment and diversity of Late Pleistocene Beringia.

Keywords: Ancient DNA; Beringia; Permafrost; Pleistocene; Phylogenetics

Introduction
The largest currently accessible collection of Late Pleistocene fossils began as a by-product of extensive placer gold mining activities near Fairbanks, Alaska, between 1930 and 1960. Gold-bearing gravels were exposed by removing some 10 – 15 m of fossil-rich permafrost silt using high-pressure water and steam. Together with the gold, thousands of long-frozen bones and teeth were exposed, and over 8 tons were collected largely by the efforts of the late Otto Geist. The majority of these specimens are currently housed in the Frick Collection at the American Museum of Natural History in New York.

Along with other large collections at the University of Alaska Museum (Fairbanks, Alaska) and the Canadian Museum of Nature (Aylmer, Quebec), where the collection is predominantly Yukon-derived, this material provides a unique opportunity to reconstruct the environment of Late Pleistocene Beringia. The skeletal remains provide a detailed record of changes in the morphology, distribution, and taxonomic composition of ecosystems, as well as important data about floral composition and regional climate.

These collections have been used extensively to identify and characterize extinct megafauna (e.g., Vereshchagin and Baryshnikov, 1982; Harington, 1985), investigate diet and ecology (e.g., Matheus, 1995), and, together with pollen records, insect remains and plant macrofossils to analyze the nature of the late Pleistocene Beringian environment (Cwynar and Ritchie, 1980; Guthrie, 1982, 1990; Colinvaux, 1996; Elias, 2000; Guthrie, 2001). Although this research has revealed much about Beringian paleoecology, major questions remain about the impact of large-scale events such as the last glacial maximum (LGM), human invasion, and the megafaunal extinction.

Ancient DNA (aDNA) techniques provide a new way to examine these issues by retrieving genetic information from preserved remains of plants and animals up to around 100,000 years old (Lindahl, 1993). These methods make it possible to follow the evolution of populations through time, recording the genetic consequences of events such as population fragmentation and large-scale climate change. In this article, a brief history of aDNA research and associated problems is followed by a discussion of how aDNA can be used to address some of the phylogenetic and phylogeographic problems associated with the LGM and the introduction of humans into North America. The unique capacity of permafrost remains to improve understanding of biomolecular decay is also discussed.

* Corresponding author. Fax: +44-1865-271249.
E-mail address: alan.cooper@zoo.ox.ac.uk (A. Cooper).
A brief history of ancient DNA

Ancient DNA initially rose to scientific attention in 1984, when 229 base pairs (bp, the units of DNA) of mitochondrial DNA (mtDNA) sequence were retrieved from a 140-year-old skin of the quagga (Equus quagga), an extinct South African relative of the zebra (Equus zebra) (Higuchi et al., 1984). The data demonstrated that the quagga was more closely related to the zebra than to other horses and revealed the approximate date of divergence between the quagga and zebra. Although not a huge taxonomic advance in itself, this result demonstrated that DNA could survive after death and heralded the possibility of using genetic data directly from the past to address evolutionary and ecological questions.

DNA is preserved in a variety of plant and animal remains, including leaf tissue, seeds, bone, hair, teeth, skin, and fecal/urine waste products. While an organism is alive, an array of enzymes serves to maintain DNA by repairing strands broken and damaged by oxidative and hydrolytic processes, free radicals, and radiation. However, after death this repair system quickly stops and damaged DNA remains unrepaired (Pääbo, 1989; Lindahl, 1993, 1997). Furthermore, as cellular processes collapse and bacteria spread, autolytic enzymes (which normally degrade foreign bodies, DNA, and unwanted products) are released from control, causing further DNA degradation. Empirical evidence suggests that the bulk of this degradation occurs rapidly after death and that these processes can be considerably slowed in certain environments—particularly cold or rapidly desiccating conditions. However, biochemical data suggest the upper limit for DNA survival is on the order of a few hundred thousand years (Lindahl, 1993).

The years following the introduction of the polymerase chain reaction (PCR) (Mullis and Faloona, 1987) saw aDNA research grow rapidly (Fig. 1). PCR is an enzyme-based iterative reaction in which the number of copies of the target sequence increases exponentially. To aDNA researchers, this means that even when only extremely small amounts of DNA survive, enough “photocopies” of the sequence can be generated to perform complex genetic analyses. However, the amplifying power of PCR also brings with it an extreme sensitivity to contamination, as even trace amounts of modern DNA will preferentially amplify and outcompete an ancient template. Consequently, reports of the amplification of chloroplast DNA from a 17- to 20-million-year-old magnolia leaf (Golenberg et al., 1990), amber-entombed insects (Desalle et al., 1992; Cano et al., 1993), and dinosaur bones (Woodward et al., 1994) were all received with considerable scepticism, particularly as attempts to replicate them failed or exposed them as modern contamination (Pääbo and Wilson, 1991; Sidow et al., 1991; Soltis, 1995; Yousten and Rippere, 1997).

Obstacles and challenges

The polymerase enzyme used in PCR to copy DNA will preferentially amplify undamaged modern DNA rather than damaged and modified aDNA even when the ancient DNA is present in higher concentrations (Handt et al., 1994). This problem becomes severe when aDNA work is performed in a building where other molecular biology research occurs, due to the massive concentrations of DNA produced by cloning and PCR. For example, a successful PCR reaction can contain around 10^{12} molecules of amplified product, and microscopic aerosol droplets created while setting up experiments can easily deposit millions of copies of the
product on researchers, equipment, and laboratory surfaces. Movement of personnel, cleaning activities, and air-conditioning will quickly distribute this throughout a building. As a consequence, aDNA research must be performed under the most stringent conditions of physical isolation and with a heavy burden of proof (Cooper and Poinar, 2000). Ancient human DNA studies are particularly suspect due to the multiple possibilities for contamination by modern humans and problems with authentication (Cooper and Poinar, 2000; Hofreiter et al., 2001a). More than 15 years after the initial aDNA reports, the field is still coming to terms with these issues.

In addition to contamination introduced within the lab, buried specimens are likely to be heavily contaminated with fungal and bacterial DNA, which provide both a background level of genetic material and possible alternative templates for PCR if the target is a microbial pathogen (Handt et al., 1994). Excavation/collection and storage may lead to substantial amounts of human DNA being deposited on exterior surfaces of the sample, and absorption also allows deep interior penetration. Curatorial activities, such as the use of animal-based glues, or storage practices in which ancient and modern specimens come into contact provide a further possible contamination risk.

The range of possible genetic targets in ancient samples is limited by DNA degradation, and consequently most research has been constrained to mitochondrial DNA (mtDNA) sequences, which are present in much higher copy numbers than nuclear DNA. Mitochondria are small energy-producing organelles present in hundreds to thousands of copies within eukaryotic cells. However, a significant problem with mtDNA is the existence of nuclear-encoded copies of mitochondrial sequences (numts) that can coamplify, effectively contaminating a PCR (Zhang and Hewitt, 1996). Numts can be amplified in addition to, or instead of, authentic aDNA if they are preferentially targeted by PCR primers or if the numt is present in the genome at high copy number, as seems to be the case for certain modern groups such as felids and elephantids (Lopez et al., 1996; Hansen et al., 2001). Although this would seem less likely with ancient samples, it has been reported in several studies (Greenwood et al., 1999; Barnes et al., 2000; Vanderkuyl et al., 1995). Numts are a serious problem for ancient DNA studies as standard methods of circumvention (e.g., long-range PCR and differential extraction of mtDNA) are not possible. Consequently, careful primer design and extensive cloning of PCR products are required.

**Necessary materials**

The amount of material necessary for DNA analysis varies according to the condition of the specimen. It is rarely necessary to use more than 1 cm$^3$ of bone (Fig. 2) or a single tooth root from permafrost specimens, although the amount of recoverable DNA can decrease rapidly if the specimens are exposed for several seasons. The best material for DNA analysis has the appearance of modern bone: solid and heavy and also dense in appearance with few visible cracks. Dark coloration can signify staining with humic acids or more significant decay, although in some conditions bone material only a millimeter beneath a deeply discolored surface may be in almost perfect condition. Unfortunately, it is difficult to determine the condition of DNA without destructive sampling. For this reason, efforts are ongoing to develop a less destructive test for the presence of DNA.

**Permafrost-preserved potential**

The constant cold of permafrost regions favors the long-term preservation of DNA, but empirical evidence suggests that the conditions immediately postmortem are a significant influence, with reworking of bones another factor. Freezing shortly after death limits the amount of DNA damage from enzymes intrinsic to the cell or subsequent bacterial or fungal degradation (Lindahl, 1993). However, Guthrie (1990) has observed that an intact carcass generates enough heat during decomposition to limit freezing, and this may explain why the DNA yield from mummified permafrost specimens is generally much lower than from isolated bones. It is possible that the best DNA samples may come from bones that have been rapidly defleshed after death, preventing bacterial growth in close proximity to the bone.

In 1994, two groups simultaneously published studies in which ancient DNA was successfully extracted from Siberian mammoth bones ranging from 9700 to more than 50,000 years old (Hagelberg et al., 1994; Höss et al., 1994). These studies were the first to demonstrate the potential of permafrost-preserved Pleistocene remains in aDNA studies. Since that time, additional mammoth sequences have been reported (e.g., Yang et al., 1996; Ozawa et al., 1997), along with those from brown bears (Talbot and Shields, 1996; Leonard et al., 2000; Barnes et al., 2002), short-faced bears (Barnes et al., 2002), and horses (Vila et al., 2001), while research is currently underway on remains of wolves, bison, moose, caribou, and lions, among others. The number of specimens (particularly of mammoth, horses, and bison) makes it possible to perform large-scale analyses of mammal populations over a considerable range of time and habitat. The areas around Fairbanks, Alaska, and Dawson City, Yukon Territory, are particularly well represented among DNA-bearing bones, as gold-mining activity has led to their discovery and collection after very limited surface exposure. Consequently, our current research focus includes considerable efforts to obtain genetic data from across the entire Beringian landmass to complement the patterns observed in the two more heavily sampled regions.
Taxonomy and systematics

As with the quagga (Higuchi et al., 1984), the first aDNA studies using permafrost remains focused on a systematic question: that of the taxonomic identity of the mammoth. Höss et al., (1994) used aDNA sequences from four Siberean mammoths to (perhaps unsurprisingly) place mammoths in a phylogenetic position closer to elephants than to other ungulates. Later, in concurrence with morphological evidence, Thomas et al., (2000) suggested mammoths were more closely related to African elephants than to Asian elephants. The DNA evidence was weakened, however, by the problems caused by the elephantid numts, which also call into question the only mastodon sequences reported to date (Yang et al., 1996; Greenwood et al., 1999).

Ancient DNA offers a means to address long-standing taxonomic problems raised by morphological variation. For example, taxonomic divisions among steppe bison have been a contentious issue, with several species designations employed to differentiate among morphologically diverse types. Part of the problem stems from a dramatic change in bison horn size and shape following the LGM (Guthrie, 1990). Our preliminary data suggest that this taxonomic arrangement is inaccurate, and that the observed rapid morphological change is due to extreme morphological plasticity rather than extinction of one population followed by replacement with a new, genetically distinct type. Permafrost bison are just one example of several ongoing megafaunal systematic studies. Others include short-faced bears (Arctodus), brown bears (Ursus arctos), and the relationships among the North American paleofelids (e.g., Homotherium, Smilodon, and Panthera).

While it is possible to use preserved genetic information to identify specimens that are incomplete or too badly damaged to classify morphologically, the cost of analyses makes this an impractical option for regular use. However, when an accurate identification is vital in determining the presence/absence of a species from a crucial site or time period, for example if a specimen is found in association with an early Native American site, aDNA can be an extremely useful tool.

Fig. 2. Examples of samples taken for DNA analysis from (a) a brown bear (Ursus arctos) ulna and (b) a bison (Bison sp.) metatarsal. Both bones are part of the Canadian Museum of Nature collection and originate from the Dawson City area, Yukon Territory.
Genetic records of environmental change

The fragmentation of formerly Holarctic large mammal populations by advancing glacial and climatic changes associated with the LGM provides an opportunity to examine the genetic effects of global climate change on ecosystems. Detailed records of the genetic history of a population provide many novel opportunities for evolutionary research. Figure 3 is an illustration of one way in which aDNA can provide insight into the phylogeographic history of a population. At some point in the past, the hypothetical population had been much more genetically diverse and widespread. Some local isolation of genetic characters had occurred, and three distinct genetic groups, called clades, were formed. The population existed both in the north and in the south; however, Clade C was unique to the northern region. The modern population is only found in the south and consists only of individuals belonging to Clade A. These data would suggest that at some point in the past, the population went through a severe bottleneck, with all of the northern individuals going extinct and only Clade A surviving in the south. If the ages of the specimens are known, it may be possible to correlate the bottleneck with a specific ecological event, such as advancing ice, a volcanic eruption, or human introduction.

The genetic records of one intensively studied group, brown bears (Leonard et al., 2000; Barnes et al., 2002), have revealed surprisingly complex patterns of localized extinctions, replacements, and changing biotic interactions during the Late Pleistocene. In this case, a number of genetic events have generated a relatively simple late Holocene phylogeographic distribution, leading to incorrect conclusions about ancient patterns of distribution. Despite a dynamic history, an unexpected degree of phylogeographic structure is observed in late Pleistocene mitochondrial haplotypes, suggesting that some form of regionalism has existed within these populations over considerable time and space. Changes to this relatively rigid phylogeographic structure enable the timing of significant environmental events to be inferred and reveal the location and timing of possible dispersal routes between Beringia and the landmasses south of the ice sheets (Leonard et al., 2000). Importantly, the timing of these late Pleistocene genetic events does not appear to be tightly related to major climatic changes and it appears, in the bears at least, that environmental effects may be secondarily mediated through interaction with competitors or other factors (Barnes et al., 2002).

Recent advances in the analysis of genetic sequences provide a method to date relatively recent evolutionary events, such as the timings of population invasions into particular locations. Bayesian statistical inference approaches using Markov Chain Monte Carlo (MCMC) simulations have been designed that allow information about the “absolute” ages of noncontemporaneous samples to provide information about the rates of DNA sequence evolution within species (Drummond and Rodrigo, 2000). The MCMC methods also allow divergence times and effective population sizes to be calculated, permitting the analysis of deep evolutionary events within the phylogenetic trees; for example, the date of separation of the polar bear and the brown bear (Heaton et al., 1996). By contrasting the genetic records of several disparate taxa, it should be possible to identify key periods of clade diversification or genetic bottlenecks, perhaps identifying periods of widespread environmental change. The data required for this powerful synthesis should be available within the next few years.

DNA preservation and diagenesis

Permafrost material also provides an important opportunity to explore the processes and temporal limits of DNA preservation in bone. Collagen levels in permafrost-preserved bone are sufficiently high that accelerator mass spectrometry radiocarbon dates have been obtained back to 58,000 $^{14}$C years B.P. using ultrafiltration—about 20,000 years older than the normal limits. This potentially enables a DNA studies to record and examine events from oxygen isotope stage 3 onwards, and indeed several DNA sequences have been retrieved from material older than this, that is, not datable by $^{14}$C methods. Furthermore, preliminary results show that large PCR amplifications are possible in some permafrost-preserved material, with amplifications as large as 2000 bp obtained from bison specimens from Siberia and the Alaskan North Slope (Shapiro and Cooper, unpublished data). This is unusual because ancient DNA fragments larger than 150–200 bp cannot normally be amplified and this raises the possibility that single-copy nuclear genes may be accessible. This potential is of great interest, because the
nuclear genome contains a vast amount of information compared to the mtDNA, allowing far more complex analyses.

Future aims

The past decade has seen great advances in the techniques used to retrieve ancient DNA from nonhuman remains (Hofreiter et al., 2001b). These genetic records allow significant biogeographic events to be detected and the effects on different taxa to be contrasted. They also provide a means to record evolution within populations over geologically significant time frames, revealing the background to the genetic distribution and diversity of their modern descendants. Preliminary data have shown the efficacy of this approach, with large data sets revealing a number of unexpected and provocative genetic patterns. The synthesis of genetic, geological, paleontological, archaeological, and paleoecological data will be essential to extracting the maximum information from this Ice Age genetic museum.

Acknowledgments

We thank Richard Tedford, John Alexander, and the American Museum of Natural History; Richard Harington, Kieran Shepherd, and the Canadian Museum of Nature; Paul Matheus, Gordon Jarrell, Dale Guthrie, Gerry Shields, Roland Gangloff, and the University of Alaska Museum; John Storer, Jeff Hunston, and the Yukon Heritage Branch; Rick Farnell, Loralee Laberge, and the Yukon Department of Renewable Resources; Beron Mining Company; and the many other museum curators, miners, and private collectors who have very kindly allowed us to take samples for genetic research. We thank Richard Harington, John Storer, and one anonymous reviewer for helpful comments on the article. The Oxford University Museum of Natural History generously provided us with lab space, and the Oxford Radiocarbon Accelerator Unit and NERC have facilitated our dating program. This work was funded by the Rhodes Trust (BS), NERC, and the Wellcome Trust (AC).

References


