Genomic Sequencing of Pleistocene Cave Bears

James P. Noonan, 1,2 Michael Hofreiter, 3 Doug Smith, 1 James R. Priest, 2 Nadim Rohland, 3 Gernot Rabeder, 4 Johannes Krause, 3 J. Chris Detter, 1,5 Svante Pääbo, 3 Edward M. Rubin 1,2*

Despite the greater information content of genomic DNA, ancient DNA studies have largely been limited to the amplification of mitochondrial sequences. Here we describe metagenomic libraries constructed with unamplified DNA extracted from skeletal remains of two 40,000-year-old extinct cave bears. Analysis of ~1 megabase of sequence from each library showed that despite significant microbial contamination, 5.8 and 1.1% of clones contained cave bear inserts, yielding 26,861 base pairs of cave bear genome sequence. Comparison of cave bear and modern bear sequences revealed the evolutionary relationship of these lineages. The metagenomic approach used here establishes the feasibility of ancient DNA genome sequencing programs.

Genomic DNA sequences from extinct species can help reveal the process of molecular evolution that produced modern genomes. However, the recovery of ancient DNA is technologically challenging, because the molecules are degraded and mixed with microbial contaminants, and individual nucleotides are often chemically damaged (1, 2). In addition, ancient remains are invariably contaminated with modern DNA, which amplifies efficiently compared with ancient DNA, and therefore inhibits the detection of ancient genomic sequences (1, 2). These factors have limited most previous studies of ancient DNA sequences to polymerase chain reaction (PCR) amplification of mitochondrial DNA (3–8). In exceptional cases, small amounts of single-copy nuclear DNA have been recovered from ancient remains less than 20,000 years old obtained from permafrost or desert environments, which are well suited to preserving ancient DNA (9–12). However, the remains of most ancient animals, including hominids, have not been found in such environments.

To circumvent these challenges, we developed an amplification-independent direct-cloning approach to constructing metagenomic libraries from ancient DNA (Fig. 1). Ancient remains are obtained from natural environments in which they have resided for thousands of years, and their extracted DNA is a mixture of genome fragments from the ancient organism and sequences derived from other organisms in the environment. A metagenomic approach, in which all genome sequences in an environment are anonymously cloned into a single library, may therefore be a powerful alternative to the targeted PCR approaches that have been used to recover ancient DNA molecules. We chose to explore this strategy with the extinct cave bear instead of an extinct hominid, to unambiguously assess the issue of modern human contamination (1, 2). In addition, because of the close evolutionary relationship of bears and dogs, cave bear sequences in these libraries can be identified and classified by comparing them to the available annotated dog genome. The phylogenetic relationship of cave bears and modern bear species has also been inferred from mitochondrial sequences, providing the opportunity to compare the phylogenetic information content of cave bear mitochondrial and genomic DNA (13).

We extracted DNA from a cave bear tooth recovered from Ochsenhaldt Cave, Austria, and a cave bear bone from Gamszulzen Cave, Austria, dated at 42,290 (error +970/−870) and 44,160 (+1400/−1190) years before the present, respectively, by accelerator mass spectrometry radiocarbon dating (table S1). We used these ancient DNA molecules to construct two metagenomic libraries, designated CB1 and CB2 (Fig. 1) (14). These libraries were constructed in a laboratory into which modern carnivore DNA has never been introduced. Ancient DNA molecules were blunt end-repaired before ligation but were otherwise neither enzymatically treated nor amplified. We sequenced 9035 clones [1.06 megabases (Mb)] from library CB1 and 4992 clones (1.03 Mb) from library CB2. The average insert sizes for each library were 118 base pairs (bp) and 207 bp, respectively.

We compared each insert in these libraries to GenBank nucleotide, protein, and environmental sequences, and the July 2004 dog whole genome shotgun assembly, by using Basic Local Alignment Search Tool (BLAST) software with an expect value cutoff of 0.001 and a minimum hit size of 30 bp (14–16). 1.1% of clones in library CB1 (Fig. 2A) and 5.8% of clones in library CB2 (Fig. 2B) had significant hits to dog genome or modern bear sequences. Our direct-cloning approach produces chimeric inserts, so we defined as candidate cave bear
sequence only that part of the insert that had a hit to dog or bear sequence. The average hit for the 100 clones in library CB1 with significant hits to carnivore sequence was 68 bp long and covered 58% of the insert; whereas the average hit for the 289 clones in library CB2 with significant carnivore hits was 70 bp and covered 49% of the insert. None of these clones showed homology to cave bear mitochondrial DNA sequences, which is consistent with the expected ratio of nuclear versus mitochondrial clones, given the much greater size of the nuclear genome (14). Based on the amount of cave bear genomic DNA used to construct library CB2, we estimate that it could yield >10-fold coverage of the cave bear genome if sequenced completely (14).

BLAST hits to the dog genome from libraries CB1 and CB2 were on average 92.4 and 92.3% identical to dog, respectively. To confirm that these sequences were indeed those of the cave bear, we designed primers against 124 putative cave bear sequences and successfully amplified and sequenced 116 orthologous sequences from the modern brown bear. All 116 of these modern bear sequences were at least 97% identical to their cave bear orthologs, verifying that all or nearly all of the carnivore sequences in both libraries are genuine cave bear genomic sequences. Only 6 of 14,027 sequenced clones had an insert that was identical to modern human genomic DNA (Fig. 2, A and B). The average BLAST hit length to the human genome for these clones was 116 bp, and the average hit covered 76% of the insert. Although we cannot establish a formal insert length or insert coverage threshold that differentiates between ancient and modern inserts because of the limited number of modern sequences we obtained, these values are significantly greater than the corresponding values for clones with cave bear sequences ($P < 0.05$ for the difference in both average clone length and average insert coverage calculated by two-tailed $t$ test). This result suggests that it may be possible to discriminate between inserts derived from short ancient DNA molecules and inserts containing modern undamaged DNA in ancient DNA libraries. This may have relevance to the application of these methods to ancient hominids, in which the ability to distinguish ancient hominid DNA from modern contamination will be essential.

The remaining inserts with BLAST hits to sequences from known taxa were derived from other eukaryotic sources, such as plants or fungi, or from prokaryotic sources (bacteria and archaea), which provided the majority of known sequences in each library. The end-repair reaction performed on each ancient DNA extract is likely biased toward less-damaged ancient DNA fragments and modern DNA, which could contribute to the abundance of prokaryotic sequences relative to cave bear sequences in these libraries. The representation we observe in the libraries thus reflects the proportion of clonable sequences from each source, not the true abundances of such sequences in the original extracts. However, the results from both libraries demonstrate that substantial quantities of genuine cave bear genomic DNA are efficiently end-repaired and cloned, despite this possible bias. A considerable fraction of inserts in each library (17.3% in library CB1 and 11.2% in library CB2; Fig. 2) had hits only to uncharacterized environmental sequences. The majority of these
clones had BLAST hits to GenBank sequences derived from a single soil sample (17), consistent with the contamination of each cave bear bone with soil bacteria from the recovery site. As in other metagenomic sequencing studies, most inserts in each library had no similarity to any sequences in the public databases.

To annotate cave bear genomic sequences, we aligned each cave bear sequence to the dog genome assembly using BLAST-like alignment tool (BLAT) (18). 6.1% of 6775 cave bear nucleotides from library CB1 and 4.1% of 20,086 cave bear nucleotides from library CB2 aligned to predicted dog RefSeq exons, in a total of 21 genes distributed throughout the dog genome (Fig. 2, C and D, and table S2). 4.1% and 6.2% of cave bear nucleotides, respectively, from library CB1 and library CB2 aligned to constrained nonexonic positions in the dog genome with phastCons conservation scores > 0.8 (conserved noncoding, Fig. 2, C and D) (14). The majority of cave bear sequence in each library, however, aligned to dog repeats or regions of the dog genome with no annotated sequence features. These latter sequences are likely fragments of neutrally evolving, nonrepetitive sequence from the cave bear genome. Constrained sequences are slightly overrepresented in our set: Only 1.7% of bases in the dog genome assembly were annotated as RefSeq exons, and ~10% of the cave bear sequences we obtained appear to be constrained overall, whereas 5 to 8% of positions in sequenced mammalian genomes are estimated to be under constraint (19). This discrepancy may be caused by our use of BLAST sequence similarity to identify cave bear sequences, an approach that is biased in favor of more-constrained sequences. Nevertheless, coding sequences, conserved non-coding sequences, and repeats appear in both cave bear genomic libraries at frequencies roughly proportional to what has been observed in modern mammalian genomes.

To determine whether the cave bear sequences we obtained contain sufficient information to reconstruct the phylogeny of cave bears and modern bears, we generated and aligned 3201 bp of orthologous sequences from cave bears and modern black, polar, and brown bears and estimated their phylogeny by maximum likelihood (Fig. 3A) (20). This phylogeny is topologically equivalent to phylogenies previously obtained using cave bear and modern bear mitochondrial DNA (13). This result further indicates that our libraries contain genuine cave bear sequences and demonstrates that we can obtain sufficient ancient sequences from those libraries to estimate the evolutionary relationships between ancient and modern lineages.

The substitution rate we estimated for cave bears is higher than that in any other bear lineage. On the basis of results from PCR-amplified ancient mitochondrial DNA, cytosines in ancient DNA can undergo deamination to uracil, which results in an excess of G-to-T and C-to-T (GC–AT) transitions (21). The inflated substitution rate in cave bears is likely caused by an excess of such events, because many of the substitutions assigned to the cave bear lineage are GC–AT transitions (Fig. 3A). These presumably damage-induced substitutions complicate phylogenetic reconstruction and the identification of functional sequence differences between extinct and modern species. However, these substitutions are not randomly distributed among all cave bear sequences but are clustered on a few clones from library CB2. Thus, two clones from library CB2 have three and four GC–AT transitions specific to cave bears, an observation that is extremely unlikely given that the occurrence of true randomly arising substitutions on cave bear clones should follow a Poisson distribution (table S3). When these two clones are excluded from the analysis (Fig. 3B), the apparent substitution rate and the excess of GC–AT transitions in cave bears are reduced, with little impact on estimates of substitution rates in the modern bear lineages. Although we cannot distinguish individual GC–AT substitutions from deamination-induced damage, these observations provide a quantitative means to identify cloned ancient DNA fragments with an excess of cytosine deamination events. It is also likely that heavily degraded ancient DNA fragments are poorly end-repaired and will therefore appear at reduced frequency in ancient DNA genomic libraries.

Although small amounts of genomic sequence have previously been obtained by amplification from <20,000-year-old remains (12), the direct cloning strategy employed here has yielded considerably more genomic DNA sequence from much older samples. Based on our results, ancient DNA sequencing programs for extinct Pleistocene species, including hominids, are feasible using a metagenomic approach. By revealing the phylogenomic terrain of recent mammalian evolution, these efforts should help identify the molecular events underlying adaptive differences among modern species.

References and Notes
14. Materials and methods are available as supporting material on Science Online.
15. The dog genome sequence generated by the Broad Institute and Agencourt Biosciences was obtained from the University of California at Santa Cruz Dog Genome Browser (http://genome.ucsc.edu).
18. W. J. Kent, Genome Res. 12, 656 (2002).
20. We cannot formally exclude alternative topologies, given the small number of nucleotide sites in the analysis and the limited divergence among these bear species, which also accounts for the low percent support for the internal nodes. Damage-induced substitutions in cave bears also perturb the phylogeny. The topology with the next-best likelihood score has cave bears as the outgroup to the modern bears due to an excess of GC–AT substitutions in several library CB2 sequences.
22. Data have been deposited into GenBank with accession
Marked Decline in Atmospheric Carbon Dioxide Concentrations During the Paleogene

Mark Pagani,1 James C. Zachos,2 Katherine H. Freeman,3 Brett Tipple,1 Stephen Bohaty2

The relation between the partial pressure of atmospheric carbon dioxide (pCO2) and Paleogene climate is poorly constrained. We used stable carbon isotopic values of di-unsaturated alkenones extracted from deep sea cores to reconstruct pCO2 from the middle Eocene to the late Oligocene (~45 to 25 million years ago). Our results demonstrate that pCO2 ranged between 1000 to 1500 parts per million by volume in the middle to late Eocene, then decreased in several steps during the Oligocene, and reached modern levels by the latest Oligocene. The fall in pCO2 likely allowed for a critical expansion of ice sheets on Antarctica and promoted conditions that forced the onset of terrestrial C4 photosynthesis.

The early Eocene (~52 to 55 million years ago (Ma)) climate was the warmest of the past 65 million years. Mean annual continental temperatures were considerably elevated relative to those of today, and high latitudes were ice-free, with polar winter temperatures ~10°C warmer than at present (1–3). After this climatic optimum, surface- and bottom-water temperatures steadily cooled over ~20 million years of (4, 5), interrupted by at least one major ephemeral warming in the late middle Eocene (6). High-latitude cooling eventually sustained small Antarctic ice sheets by the late Eocene (7), culminating in a striking climate shift across the Eocene/Oligocene boundary (E/O) at 33.7 Ma. The E/O climate transition, Earth’s first clear step into “icehouse” conditions during the Cenozoic, is associated with a rapid expansion of large continental ice sheets on Antarctica (8, 9) in less than ~350,000 years (10, 11).

Changes in the partial pressure of atmospheric carbon dioxide (pCO2) are largely credited for the evolution of global climates during the Cenozoic (12–14). However, the relation between pCO2 and the extraordinary climatic history of the Paleogene is poorly constrained. Initial attempts to estimate early Paleogene pCO2 have provided conflicting results, with both high (15) and low (i.e., similar to modern) (16) estimates of pCO2. This deficiency in our understanding of the history of pCO2 is critical, because the role of CO2 in forcing long-term climate change during some intervals of Earth’s history is equivocal. For example, Miocene pCO2 records (~25 to 5 Ma) argue for a decoupling between global climate and CO2 (15–17). These records suggest that Miocene pCO2 was rather low and invariant across periods of both inferred global warming and high-latitude cooling (17). Clearly, a more complete understanding of the relation between pCO2 and climate change requires the extension of paleo-pCO2 records back into periods when Earth was substantially warmer and ice-free.

Paleoatmospheric CO2 concentrations can be estimated from the stable carbon isotopic compositions of sedimentary organic molecules known as alkenones. Alkenones are long-chained (C37–C39) unsaturated ethyl and methyl ketones produced by a few species of Haptophyte algae in the modern ocean (18). Alkenone-based pCO2 estimates derive from records of the carbon isotopic fractionation that occurred during marine photosynthetic carbon fixation (εp). Chemosat experiments conducted under nitrate-limited conditions indicate that alkenone-based εp values (εp37:2) vary as a function of the concentration of aqueous CO2 ([CO2aq]) and specific growth rate (19–21). These experiments also provide evidence that cell geometry accounts for differences in εp among marine microalgae cultured under similar conditions (21). In contrast, results from dilute batch cultures conducted under nutrient-replete conditions yield substantially lower εp37:2 values, a different relation for εp versus μ/CO2 (where μ = algal growth rate), and a minimal response to [CO2aq] (22). Thus, comparison of the available culture data suggests that different growth and environmental conditions potentially trigger different carbon uptake pathways and carbon isotopic responses (23). A recent evaluation of the efficacy of the alkenone-CO2 approach, using sedimentary alkenones in the natural environment, supported the capacity of the technique to resolve relatively small differences in water column [CO2aq] across a variety of marine environments when phosphate concentrations and temperatures are constrained (24).

In our study, we extended records of the carbon isotopic composition of sedimentary alkenones (δ13C37:2) from the middle Eocene to the late Oligocene and established a record of pCO2 for the past ~45 million years. Samples from Deep Sea Drilling Project sites 516, 511, 513, and 612 and Ocean Drilling Program site 803 (Fig. 1) were used to reconstruct δ13C37:2 and εp37:2 records ranging from the middle Eocene to the late Oligocene (~25 to 45 Ma). These sites presumably represent a range of oceanic environments with a variety of surface-water nutrient and algal-growth conditions and thus reflect a set of environmental and physiological factors affecting both δ13C37:2 and εp37:2 values.

These data are presented as a composite record, in large part because the measurable concentration of di-unsaturated alkenones varied both spatially and temporally. Moreover, continuous alkenone records spanning the entire Eocene and Oligocene from individual sites were not recovered. As a consequence, most of the Oligocene record is represented at site 516, whereas the majority of the Eocene is represented at site 612 (Fig. 2A). Age models for each site were developed by linearly interpolating between biostratigraphic datums (25–31), calibrated to the Geomagnetic Polar Time Scale (32).

Eocene δ13C37:2 values range from ~30 to ~35 per mil (%), with the most negative values (sites 511 and 513) occurring near the E/O boundary. δ13C37:2 values increase substantially through the Oligocene with maximum values of ~27% by ~25.5 Ma. This trend is briefly reversed near the end of the Oligocene as δ13C37:2 values become more negative, reaching ~32% by 25 Ma (Fig. 2A). The overall pattern of 13C enrichment continues into the Miocene, establishing a clear secular trend from the middle Eocene to the middle Miocene (Fig. 2B). These isotopic
Genomic Sequencing of Pleistocene Cave Bears
James P. Noonan et al.
Science 309, 597 (2005);
DOI: 10.1126/science.1113485

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of March 24, 2016):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
/content/309/5734/597.full.html

Supporting Online Material can be found at:
/content/suppl/2005/07/21/1113485.DC1.html

A list of selected additional articles on the Science Web sites related to this article can be found at:
/content/309/5734/597.full.html#related

This article cites 18 articles, 12 of which can be accessed free:
/content/309/5734/597.full.html#ref-list-1

This article has been cited by 77 article(s) on the ISI Web of Science

This article has been cited by 36 articles hosted by HighWire Press; see:
/content/309/5734/597.full.html#related-urls

This article appears in the following subject collections:
Paleontology
/cgi/collection/paleo