EXTENDED PDF FORMAT
SPONSORED BY

BIO-RAD

http://discover.bio-rad.com

Sequencing and Analysis of Neanderthal Genomic DNA

James P. Noonan, *et al.*

Science **314**, 1113 (2006);
DOI: 10.1126/science.1131412

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 9, 2007):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/314/5802/1113

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/314/5802/1113/DC1

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
http://www.sciencemag.org/cgi/content/full/314/5802/1113#related-content

This article **cites 22 articles**, 10 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/314/5802/1113#otherarticles

This article has been **cited by** 5 article(s) on the ISI Web of Science.

This article appears in the following **subject collections**:
Evolution
http://www.sciencemag.org/cgi/collection/evolution

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl

Downloaded from www.sciencemag.org on April 9, 2007

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright c 2006 by the American Association for the Advancement of Science; all rights reserved. The title SCIENCE is a registered trademark of AAAS.

Sequencing and Analysis of Neanderthal Genomic DNA

James P. Noonan,[1,2] Graham Coop,[3] Sridhar Kudaravalli,[3] Doug Smith,[1] Johannes Krause,[4] Joe Alessi,[1] Feng Chen,[1] Darren Platt,[1] Svante Pääbo,[4] Jonathan K. Pritchard,[3] Edward M. Rubin[1,2]*

Our knowledge of Neanderthals is based on a limited number of remains and artifacts from which we must make inferences about their biology, behavior, and relationship to ourselves. Here, we describe the characterization of these extinct hominids from a new perspective, based on the development of a Neanderthal metagenomic library and its high-throughput sequencing and analysis. Several lines of evidence indicate that the 65,250 base pairs of hominid sequence so far identified in the library are of Neanderthal origin, the strongest being the ascertainment of sequence identities between Neanderthal and chimpanzee at sites where the human genomic sequence is different. These results enabled us to calculate the human-Neanderthal divergence time based on multiple randomly distributed autosomal loci. Our analyses suggest that on average the Neanderthal genomic sequence we obtained and the reference human genome sequence share a most recent common ancestor ~706,000 years ago, and that the human and Neanderthal ancestral populations split ~370,000 years ago, before the emergence of anatomically modern humans. Our finding that the Neanderthal and human genomes are at least 99.5% identical led us to develop and successfully implement a targeted method for recovering specific ancient DNA sequences from metagenomic libraries. This initial analysis of the Neanderthal genome advances our understanding of the evolutionary relationship of *Homo sapiens* and *Homo neanderthalensis* and signifies the dawn of Neanderthal genomics.

Neanderthals are the closest hominid relatives of modern humans (*1*). As late as 30,000 years ago, humans and Neanderthals coexisted in Europe and western Asia (*2*). Since that time, our species has spread across Earth, far surpassing any previous hominid or primate species in numbers, technological development, and environmental impact, while Neanderthals have vanished. Molecular studies of Neanderthals have been exclusively constrained to the comparison of human and polymerase chain reaction (PCR)–amplified Neanderthal mitochondrial sequences, which suggest that the most recent common ancestor of humans and Neanderthals existed ~500,000 years ago, well before the emergence of modern humans (*3*–*5*). Further analyses of mitochondrial data, including the comparison of mitochondrial sequences obtained from several Neanderthals and early modern humans, suggest little or no admixture between Neanderthal and modern human populations in Europe (*3*, *4*, *6*, *7*). However, a major limitation of all prior molecular studies of Neanderthals is that mitochondrial sequences reflect only maternal inheritance of a single locus. Accordingly, in the absence of Neanderthal autosomal and Y-chromosome sequences, the assessment of human-Neanderthal admixture remains incomplete. Mitochondrial data also provide no access to the gene and gene regulatory sequence differences between humans and Neanderthals that would help to reveal biological features unique to each. These insights await the recovery of Neanderthal genomic sequences.

The introduction of high-throughput sequencing technologies and recent advances in metagenomic analysis of complex DNA mixtures now provide a strategy to recover genomic sequences from ancient remains (*8*–*11*). In contrast to previous efforts to obtain ancient sequences by direct analysis of extracts (*3*–*6*, *12*), metagenomic libraries allow the immortalization of DNA isolated from precious ancient samples, obviating the need for repeated destructive extractions (*10*). In addition, once an ancient DNA fragment is cloned into a metagenomic library, it can be distinguished from contamination that might be introduced during subsequent PCR amplification or sequencing by the vector sequences linked to each library-derived insert (Fig. 1).

Recovery of Neanderthal nuclear DNA sequences using a metagenomic approach. In this study, we applied an amplification-independent direct cloning method to construct a Neanderthal metagenomic library, designated NE1, using DNA extracted from a 38,000-year-old specimen from Vindija, Croatia (*6*, *13*). We have recovered 65,250 base pairs (bp) of Neanderthal genome sequence from this library through a combination of Sanger sequencing and massively parallel pyrosequencing. We have also used the metagenomic library as a substrate to isolate specific Neanderthal sequences by direct genomic selection. Several lines of evidence indicated that the hominid sequences in this library were largely Neanderthal, rather than modern human contamination. Mitochondrial PCR analysis of the extract used to build the library, using an amplicon of similar size as the average hominid sequence identified in the library, revealed that only ~2% of the products were from contaminating modern human DNA, whereas the remaining 98% were Neanderthal. Signatures of damage in the hominid sequences that are characteristic of ancient DNA also suggested that they were ancient. Finally and most importantly, comparison of hominid sequences from the library to orthologous human and chimpanzee genomic sequences identified human-specific substitutions at sites where the hominid sequence was identical to that of the chimpanzee, enabling us to make estimates of the human-Neanderthal divergence time (*3*, *4*, *6*).

We initially assessed the Neanderthal genomic sequence content of library NE1 by Sanger sequencing of individual clones, which allowed individual library inserts to be completely sequenced and thus provided a direct measure of hominid insert size that could not be obtained from the ~100-bp pyrosequencing reads described below (Table 1). We identified hominid sequences in the library by

[1]U.S. Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. [2]Genomics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA. [3]Department of Human Genetics, University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA. [4]Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103, Leipzig, Germany.

*To whom correspondence should be addressed. E-mail: emrubin@lbl.gov

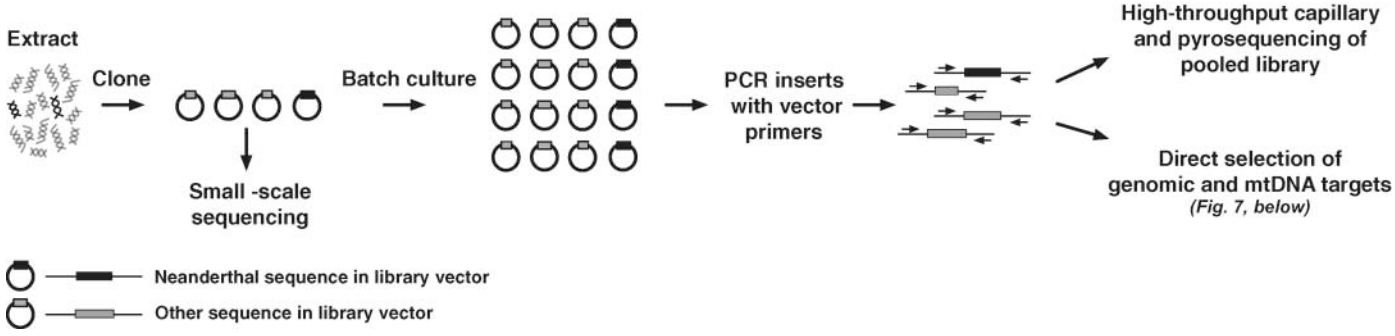

Fig. 1. Generation of ancient metagenomic library DNAs for direct selection and pyrosequencing.

Downloaded from www.sciencemag.org on April 9, 2007

BLAST comparison to the reference human genome sequence (*13*, *14*). In many cases, the human BLAST hit covered only part of the insert, because the direct cloning method we employed produces chimeric inserts consisting of smaller fragments ligated into larger concatemers. The small average size of these putatively ancient Neanderthal fragments (52 bp) is similar to results we previously obtained from two Pleistocene cave bear libraries, in which the average library insert size was between 100 and 200 bp, whereas BLAST hits to reference carnivore genome sequences were on average 69 bp (Fig. 2) (*10*). The small BLAST hit sizes and insert sizes in both cave bear and Neanderthal metagenomic libraries are consistent with the degradation of ancient genomic DNA into small fragments over tens of thousands of years, illustrating the general condition of nuclear DNA in ancient remains.

Sanger sequencing of individual clones from library NE1 suggested that it contained sufficient amounts of Neanderthal sequence to conduct a random sequence survey of the Neanderthal genome. However, the small percentage of clones we identified as containing hominid sequences indicated that we would have to sequence a very large number of clones to obtain enough Neanderthal genome sequence for this analysis. We therefore carried out deep sequencing of pooled inserts from library NE1 using massively parallel pyrosequencing. To obtain pooled inserts, we amplified transformed NE1 library DNA in liquid batch culture and recovered library inserts from purified plasmid DNA by PCR (Fig. 1). We generated 1.47 million pyrosequencing reads, compared each to the human genome sequence with MEGABLAST, and obtained 7880 hits. Assembly of these reads and reanalysis of the resulting scaffolds by BLASTN produced 1126 unique Neanderthal loci, yielding 54,302 bp of Neanderthal genomic sequence (*13*).

Assessment of pyrosequencing data quality by comparison to Sanger sequence data. The pyrosequencing approach generates significant amounts of sequence but does so with a higher error rate than Sanger sequencing (*11*). To assess the quality of Neanderthal pyrosequencing data, we generated consensus sequences from pyrosequencing reads overlapping the same Neanderthal genomic locus and filtered out low-quality positions in the resulting contigs (quality score < 15). To determine whether these contigs contained additional errors not detectable by quality-score filtering, we also used Sanger sequencing to analyze 19,200 clones from the same batch culture used to generate the pyrosequencing data. This sequencing yielded 130 loci (6.2 kb) that were also represented in the pyrosequencing data. Sanger sequencing and pyrosequencing results for these 130 Neanderthal loci agreed at 99.89% of ungapped positions. In addition, Sanger sequencing and pyrosequencing yielded Neanderthal sequences that were nearly equally divergent from the human reference sequence (pyrosequencing = 0.47% divergence, Sanger sequencing = 0.49%). These results indicate that the frequency of single-base errors is probably no greater in Neanderthal genomic sequence obtained from assembled quality-filtered pyrosequencing data than in sequence obtained from Sanger sequencing.

The low complexity of library NE1 made these analyses possible, because it resulted in a limited number of clones in the library that were amplified by batch culture and PCR and then sequenced in depth (fig. S1). We estimated that the coverage obtained in library NE1 (~0.002%) is significantly lower than that previously obtained in cave bear metagenomic libraries prepared from samples of similar age as the Neanderthal sample used here (*10*). The low coverage in library NE1 is more likely due to the quality of this particular library rather than being a general feature of ancient DNA. Nevertheless, we were able to obtain substantial amounts of authentic Neanderthal genomic sequence from the library by deep sequencing.

Comparison of orthologous Neanderthal, human, and chimpanzee genomic sequences. To ascertain whether the library NE1 hominid sequence we obtained was a representative sampling of the Neanderthal genome, we identified each NE1 library sequence for which the bit score of the best BLASTN hit in the human genome was higher than the bit scores of all other hits for that sequence. We then determined the distribution of all such best BLASTN hits across human chromosomes [43,946 bp in 1,039 loci (table S1 and Fig. 3A)]. The amount of Neanderthal sequence aligned to each human chromosome was highly correlated with sequenced chromosome length, indicating that the Neanderthal sequences we obtained were randomly drawn from all chromosomes (Pearson correlation coefficient = 0.904, Fig. 3A). The hominid hits included Y-chromosome sequences, demonstrating that our sample was derived from a Neanderthal male. We annotated each Neanderthal locus according to the annotations (known genes, conserved noncoding sequences, and repeats) associated with the aligned human sequence (table S2). Neanderthal sequences obtained by both Sanger sequencing and pyrosequencing showed a distribution of sequence features consistent with the known distribution of these features in the human genome (Fig. 3B). These sequences are therefore likely to represent a random sampling of the Neanderthal genome.

Comparison of authentic Neanderthal sequence with orthologous human and chimpanzee genomic sequences will reveal sites at which Neanderthal is identical to chimpanzee but at which the human sequence has undergone a mutation since the human-Neanderthal divergence. Determining the number of human-specific mutations is critical to dating the human-Neanderthal split. To identify these events, we constructed alignments of orthologous human, Neanderthal, and chimpanzee sequences and identified mutations specific to each lineage by parsimony (*15*). We identified 34 human-specific substitutions in 37,636 human, Neanderthal, and chimpanzee aligned positions, including substitutions on chromosomes X and Y that were not considered in subsequent analyses.

We also identified 171 sites with Neanderthal-specific substitutions relative to human and chimpanzee. It has been shown that nucleotides in genuine ancient DNA are occasionally chemically damaged, most frequently because of the deamination of cytosine to uracil, resulting in the incorporation of incorrect bases during PCR and sequencing (*16*). This results in an apparent excess of C-to-T and G-to-A mismatches (which are equivalent events) between the ancient sequence and the modern genomic reference sequence. We observe a significant excess of C-to-T and G-to-A mismatches (relative to T-to-C and A-to-G mismatches) between human and NE1 hominid sequences obtained by both Sanger sequencing and pyrosequencing [$P \ll 0.0005$, Fisher's exact test (Fig. 4 and table S3)]. This accounts for the large number of Neanderthal-specific substitutions we observe and further supports the supposition that the hominid sequences are Neanderthal in origin. Despite the bias toward C-to-T and G-to-A events in Neanderthal genomic sequence, the overall frequency of these events is

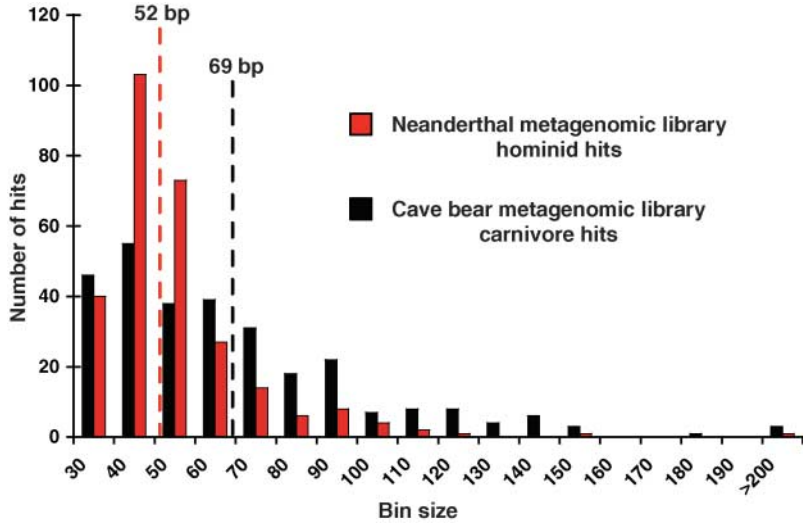

Fig. 2. Size distribution, plotted in 10-bp bins, of Neanderthal and cave bear sequences obtained from metagenomic libraries by Sanger sequencing of individual clones. The average hit size in each case is indicated by a dotted line.

Downloaded from www.sciencemag.org on April 9, 2007

low (~0.37% of all sites), indicating that the vast majority of human-Neanderthal-chimpanzee aligned positions are not likely to be significantly affected by misincorporation errors (*13*).

The length distribution of ancient DNA fragments shown in Fig. 2, when combined with the sequence signatures of ancient DNA described above, offers another metric for assessing the degree of modern human contamination in our library. Based on the assumption that modern contaminating DNA fragments would be longer than authentic ancient DNAs, which is supported by the observation that contaminating modern human DNA fragments in the cave bear libraries were on average much longer than the cave bear sequences (116 versus 69 bp) (*10*), we examined the distribution of human-Neanderthal mismatches in our data set as a function of alignment length. If a substantial fraction of the hominid sequence recovered from the Neanderthal sample were actually modern human DNA, we would expect to see a lower human-Neanderthal sequence divergence in the longer BLASTN alignments than we observe in the entire data set, because the longer hominid sequences would be enriched in modern human contaminants. The excess of damage-induced Neanderthal-specific mismatches described above would also be expected to decrease as alignment length increases, because individual bases in the longer modern human fragments would show relatively few chemical modifications. However, we did not observe these trends in our Neanderthal sequence. The human-Neanderthal sequence divergence in all autosomal alignments greater than 52 bp (the approximate midpoint of the distribution shown in Fig. 2) was similar to the divergence obtained from the whole data set (0.59% versus 0.52%). The excess of C-to-T and G-to-A mismatches was also maintained in the longer alignments. These results further support the supposition that the hominid sequence we obtained is predominantly Neanderthal in origin.

Coalescence time of human and Neanderthal genomic sequences. These data allowed us to examine for the first time the genetic relationship between humans and Neanderthals using nuclear genomic sequences (*13*). We first considered the average coalescence time for the autosomes between the Neanderthal genomic sequence that we obtained and the reference human genome sequence. We observed 502 human-chimpanzee autosomal differences in the human-Neanderthal-chimpanzee sequence alignments we constructed. Based on comparison to the Neanderthal sequence, 27 of these differences were human-specific and therefore postdate the most recent common ancestor (MRCA) of the human and Neanderthal sequences. Using this information, our maximum likelihood estimate of the average time to the MRCA of these sequences is 706,000 years, with a 95% confidence interval (CI) of 468,000 to 1,015,000 years (Figs. 5A and 6) (*13*). This calculation does not make use of Neanderthal-specific changes, because many of those events are due to DNA damage as described above. In addition, we restricted our analysis to autosomal data, because these represent 97% of our total data set and population genetic parameters are likely to differ between the autosomes and sex chromosomes. Our estimate uses a mutation rate obtained by setting the average coalescence time for human and chimpanzee autosomes to 6.5 million years ago, a value that falls within the range suggested by recent studies (*17*, *18*). Inaccuracies in the human-chimpanzee divergence time would shift all the time estimates and CIs presented here in proportion to the error.

Split time of ancestral human and Neanderthal populations. Our estimate of the average common ancestor time reflects the average time at which the Neanderthal and human reference sequences began to diverge in the common ancestral population, not the actual split time of the ancestral populations that gave rise to Neanderthals and modern humans. To estimate the actual split time of the ancestral human and Neanderthal populations, we developed a method that incorporated data from the human and Neanderthal reference sequences, as well as genotypes from 210 individuals with genome-wide single-nucleotide polymorphism (SNP) data collected by the International HapMap Consortium (Table 2) (*19*). We included the HapMap data because they indicate what proportion of sites in the Neanderthal sequence fall within the spectrum of modern human variation. For example, if the ancestral human and Neanderthal populations split long ago, before the rise of most modern human genetic diversity captured by the HapMap data, then Neanderthal sequence would almost never carry the derived allele, relative to the orthologous chimpanzee sequence, for a human SNP (Table 2). Conversely, a more recent population split would result in Neanderthal sequence frequently carrying the derived allele for human SNPs.

To formalize this idea, we considered an explicit population model for the relationship between Neanderthals and each HapMap population (East Asians,

Table 1. Amount of unique Neanderthal sequence obtained from library NE1 by Sanger sequencing of individual clones, as well as Sanger sequencing and pyrosequencing of clones in batch culture. n.a., not applicable.

	Individual clones	Batch culture	
Sequencing chemistry	Sanger	Sanger	Pyrosequencing
Reads	9984	19,200	1,474,910
Average insert	134 bp	196 bp	n.a.
Average BLAST hit	52 bp	52 bp	48 bp
Unique loci	131	69	1126
Total unique hominid sequence	6845 bp	4103 bp	54,302 bp

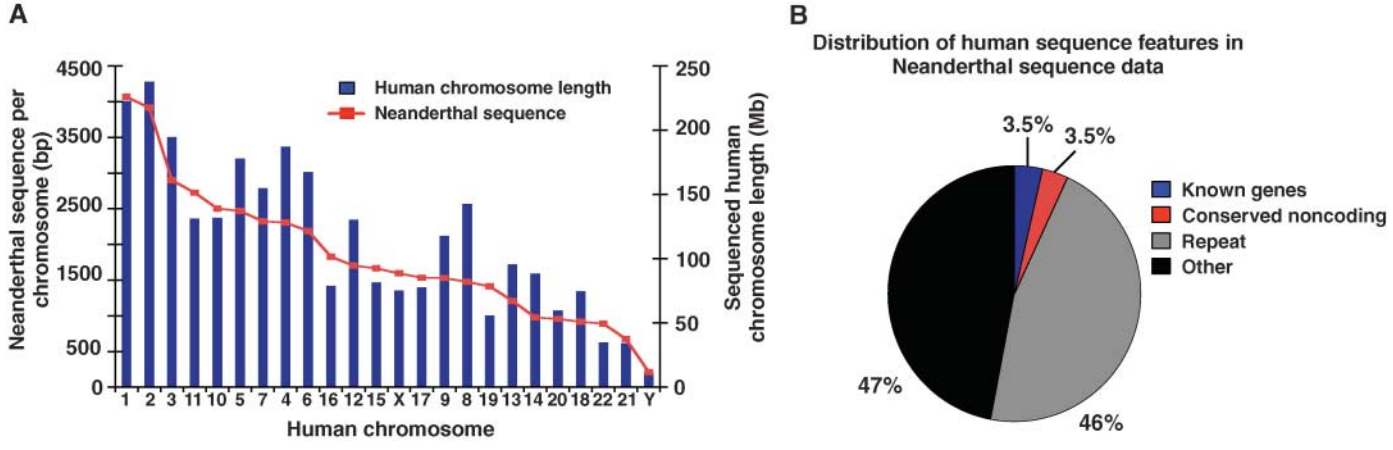

Fig. 3. **(A)** Representation of each Neanderthal chromosome in 43.9 kb of NE1 hominid sequences displaying a statistically unambiguous best BLAST hit to the human genome, relative to the total sequenced length of each human chromosome minus gaps. Chromosomes are ranked by the amount of Neanderthal sequence aligned to each. Chromosomes X and Y are shown at half their total length to correct for their haploid state in males relative to the autosomes. **(B)** Representation of sequence features in the NE1 hominid sequence shown in (A).

Downloaded from www.sciencemag.org on April 9, 2007

Europeans, and Yoruba) separately (fig. S3) (*13*). We assumed that Neanderthals and modern humans evolved from a single ancestral population of 10,000 individuals and that the Neanderthal population split away from the human ancestral population instantaneously at a time T in the past, with no subsequent gene flow. In order to model the demographic histories of the HapMap populations, we made use of models and parameters estimated by Voight *et al.* (*20*) based on resequencing data from 50 unlinked, noncoding regions. Those demographic models include bottlenecks for East Asians and Europeans and modest exponential growth for Yoruba (*13*).

We then constructed a simulation-based composite likelihood framework to estimate the time of the human-Neanderthal population split (*13*, *21*). At each site in the human-Neanderthal-chimpanzee alignments we constructed, we recorded the Neanderthal and human reference alleles relative to chimpanzee. We also determined, separately for each population, whether each site was a HapMap SNP in that population and if so, the allele frequency (Table 2).

We used simulations to estimate the probability of each possible data configuration at a single site as a function of the human-Neanderthal split time. The simulations used the estimated population demography for each HapMap population and a probabilistic model of SNP ascertainment to match the overall density and frequency spectrum of HapMap Phase II SNPs. Likelihood curves for the split time were computed by multiplying likelihoods across sites as though they were independent. In practice, this is an excellent approximation for our data because the Neanderthal sequence reads are very short and just 1 out of 905 aligned fragments contains more than one human-specific allele or SNP. Bootstrap simulations confirmed that our composite likelihood method yields appropriate CIs for the split time (*13*).

Using this approach, the maximum likelihood estimates for the split time of the ancestral human and Neanderthal populations are 440,000 years (95% CI of 170,000 to 620,000 years) based on the European data, 390,000 years (170,000 to 670,000 years) for East Asians, and 290,000 years (120,000 to 570,000 years) for Yoruba (Figs. 5B and 6). These values predate the earliest known appearance of anatomically modern humans in Africa ~195,000 years ago (*22*). Because these split times are before the migration of modern humans out of Africa, the three population-specific estimates should

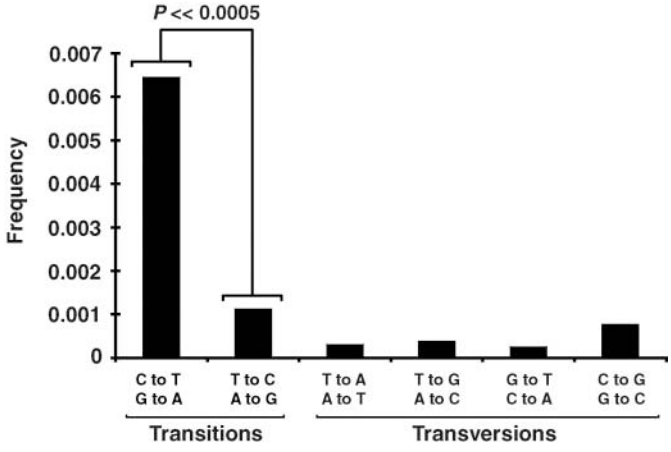

Fig. 4. Frequency distribution of 171 Neanderthal-specific substitutions observed in 37,636 bp of aligned human, Neanderthal, and chimpanzee genomic sequence. Complementary substitutions (such as C to T and G to A) are considered equivalent events.

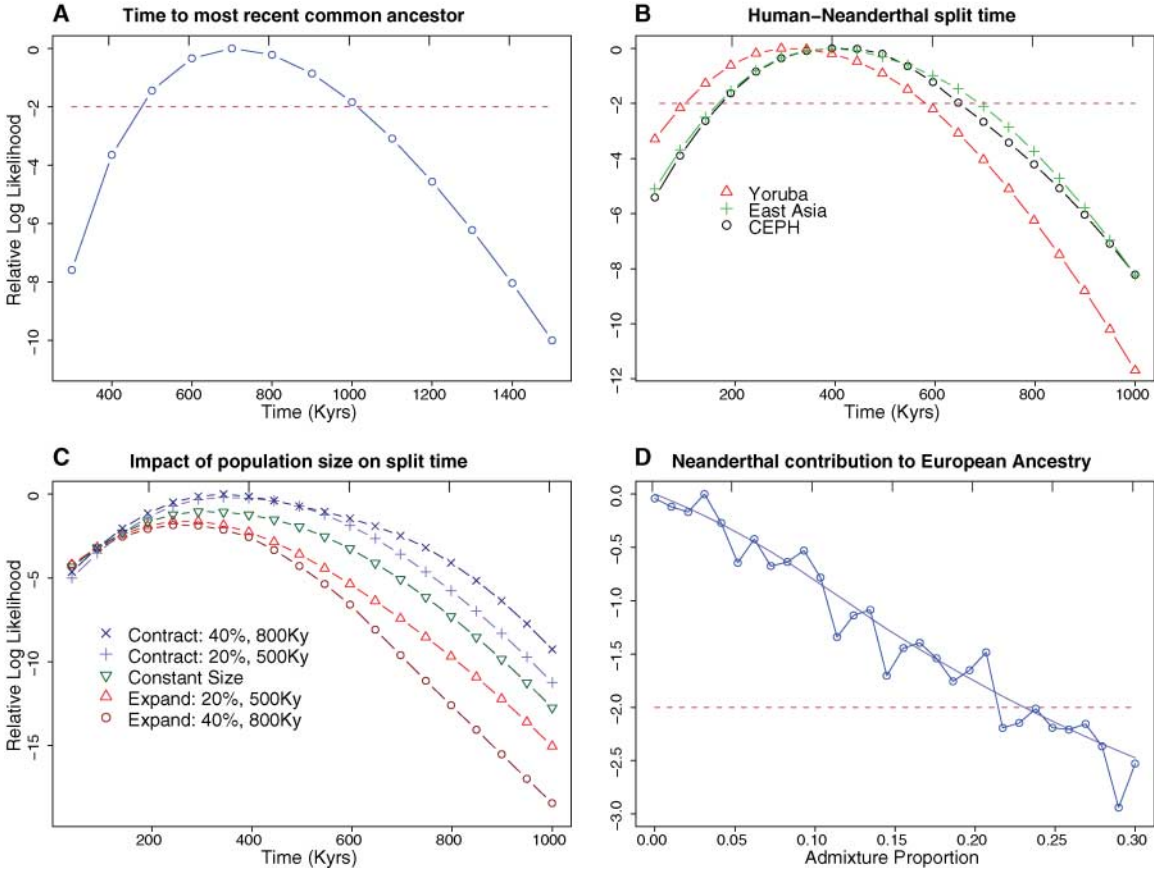

Fig. 5. (**A**) Log-likelihood curve of the time to the MRCA of the Neanderthal and human reference sequences. (**B**) Smoothed relative log-likelihood estimates of the split times between different human populations and the Neanderthal population. (**C**) Impact of changes in the ancient population size on split time estimates for five models that are consistent with modern polymorphism data. Ky, thousand years. Each curve is the smoothed log likelihood relative to the maximum over all five models. For each model, the text on the plot indicates the degree of expansion or contraction and the time before the present at which the size change occurred. The expansion models are less likely as compared to either constant population size or the contraction models. (**D**) The log-likelihood estimates of the contribution of the Neanderthal population to the ancestry of Europeans. The light blue line is a smoothed version of the estimates. The horizontal dashed maroon line in (A), (B), and (D) represents a 2 log-likelihood drop, and the region bounded by this line represents the 95% CI around the maximum likelihood estimates.

Downloaded from www.sciencemag.org on April 9, 2007

all be estimates of the same actual split time. The average of these estimates, ~370,000 years, is thus a sensible point estimate for the split time. Substantial contamination with modern human DNA would cause these estimates to be artificially low, but 2% contamination, the rate suggested by mitochondrial PCR analysis of the primary extract used to construct the library, would have essentially no impact (*13*).

Our estimates of the human-Neanderthal split time might depend heavily on the assumption that the ancestral effective population size of humans was 10,000 individuals. To address this, we explored a set of models in which the ancestral human population expanded or contracted at least 200,000 years ago (*13*). We found that much of the parameter space—though not the original model—could be excluded on the basis of modern human polymorphism data from Voight *et al*. (*20*). We repeated our likelihood analysis of the Neanderthal data using models incorporating ancient expansion or contraction that are consistent with modern data and found that these did not substantially change our population split time estimates (Fig. 5C).

Our data include three sites at which Neanderthal carries the derived allele for a polymorphic HapMap SNP. These sites are unlikely to represent modern contamination because for two of the SNPs, the derived allele is found only in Yoruba; also, one of the SNPs lies on a fragment that contains a C-to-T transition in Neanderthals that is characteristic of chemical damage to DNA. These observations indicate that the Neanderthal sequence may often coalesce within the human ancestral tree. Based on simulations of our best-fit model for Yoruba, we estimate that Neanderthal is a true outgroup for approximately 14% (assuming a split time of 290,000 years, the Yoruba estimate) to 26% (assuming a split time of 440,000 years, the European estimate) of the autosomal genome of modern humans, although more data will be required to achieve a precise estimate.

Lack of evidence for admixture between humans and Neanderthals. Because Neanderthals coexisted with modern humans in Europe, there has long been interest in whether Neanderthals might have contributed to the European gene pool. Previous studies comparing human and Neanderthal mitochondrial sequences did not find evidence of a Neanderthal genetic contribution to modern humans. However, the utility of mitochondrial data in addressing this question is limited in that it is restricted to a single locus and, due to the maternal inheritance of mitochondrial DNA, is informative only about admixture between Neanderthal females and modern human males (*3*–*6*). Moreover, it has been argued that some aspects of modern human autosomal data may be the result of modest levels of Neanderthal admixture (*23*).

If Neanderthal admixture did indeed occur, then this could manifest in our data as an abundance of low-frequency derived alleles in Europeans where the derived allele matches Neanderthal. No site in the data set appears to be of this type. In order to formally evaluate this hypothesis, we extended our composite likelihood simulations to include a single admixture event 40,000 years ago in which a fraction p of the European gene pool was derived from Neanderthals. We fixed the human-Neanderthal split at 440,000 years ago (the split time estimate for Europeans). With these assumptions, the maximum likelihood estimate for the Neanderthal contribution to modern genetic diversity is zero. However, the 95% CI for this estimate ranges from 0 to 20%, so a definitive answer to the admixture question will require additional Neanderthal sequence data (Fig. 5D).

Targeted recovery of specific Neanderthal sequences by direct genomic selection. Although we have recovered significant amounts of Neanderthal genome sequence using a metagenomic approach, hundreds of gigabases of sequence would be required to achieve reasonable coverage of a single Neanderthal genome by this method. Moreover, our results indicate that at least 99.5% of the Neanderthal sequence that would be obtained would be identical to the modern human sequence. The human-Neanderthal sequence differences that would yield great insight into human biology and evolution are thus rare events in an overwhelming background of uninformative sequence. We therefore explored the potential of metagenomic libraries to serve as substrates to recover specific Neanderthal sequences of interest by targeted methods. To this end, we developed a direct genomic selection approach to recover known and unknown sequences from metagenomic ancient DNA libraries (Fig. 7) (*24*). We first attempted to recover specific sequences from a Pleistocene cave bear metagenomic library we previously constructed. We designed PCR probes corresponding to 96 sequences highly conserved among mammals but not previously shown to be present in the cave bear library. We amplified these sequences from the human genome and hybridized the resulting probes to PCR-amplified cave bear library inserts produced as described above (Fig. 1). Recovered library DNAs were amplified by PCR and sequenced. We successfully recovered five targets consisting of a known enhancer of *Sox9* and conserved sequences near *Tbx3*, *Shh*, *Msx2*, and *Gdf6* (table S4). In principle, these sequences could be derived from contaminating DNA rather than the cave bear library. Critically, the captured cave bear sequences were flanked by library vector sequence, directly demonstrating that these sequences were derived from a cloned library insert and not from contaminating DNA introduced during direct selection (Fig. 7 and fig. S2).

Based on these results, we attempted to recover specific Neanderthal sequences from library NE1. We focused on recovering sequences that we had previously identified by shotgun sequencing because of the low complexity of library NE1, and were able to recover 29 of 35 sequences we targeted (table S4). The authenticity of these sequences was confirmed by the presence of library vector sequences in the reads. Our

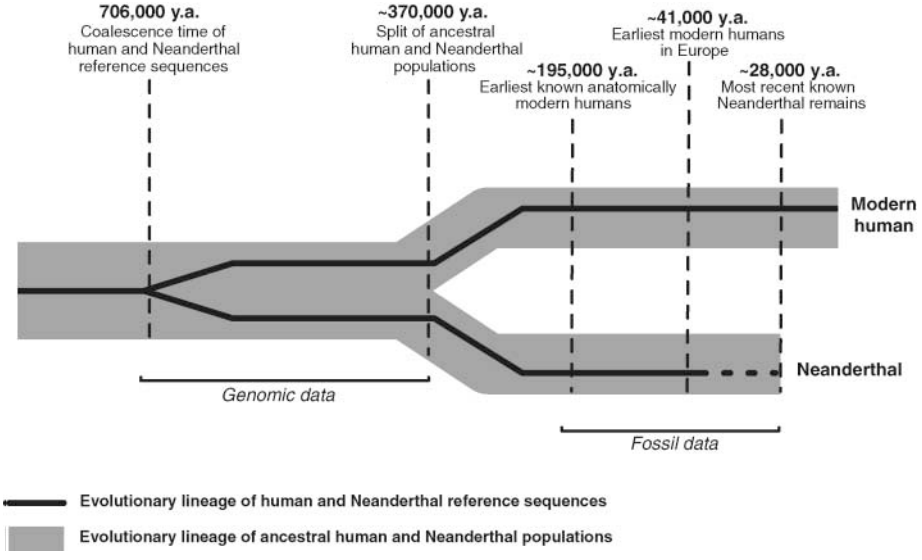

Fig. 6. Divergence estimates for human and Neanderthal genomic sequences and ancestral human and Neanderthal populations, shown relative to dates of critical events in modern human and Neanderthal evolution (*2*, *22*, *25*). The branch lengths are schematic and not to scale. y.a., years ago.

Table 2. Summary of all autosomal sites sequenced in Neanderthal and uniquely aligned to the human and chimpanzee reference sequences. The designations "ancestral" and "derived" indicate whether each site is, respectively, a match or mismatch with chimpanzee. Sites are partitioned into those that overlap a Phase II HapMap SNP (with SNPs) and those that do not (without SNPs).

		Sequence state in human reference	
	With SNPs	Ancestral	Derived
Sequence state in Neanderthal	Ancestral	24	8
	Derived	3	0
		Sequence state in human reference	
	Without SNPs	Ancestral	Derived
Sequence state in Neanderthal	Ancestral	35,801	19
	Derived	161	475

Downloaded from www.sciencemag.org on April 9, 2007

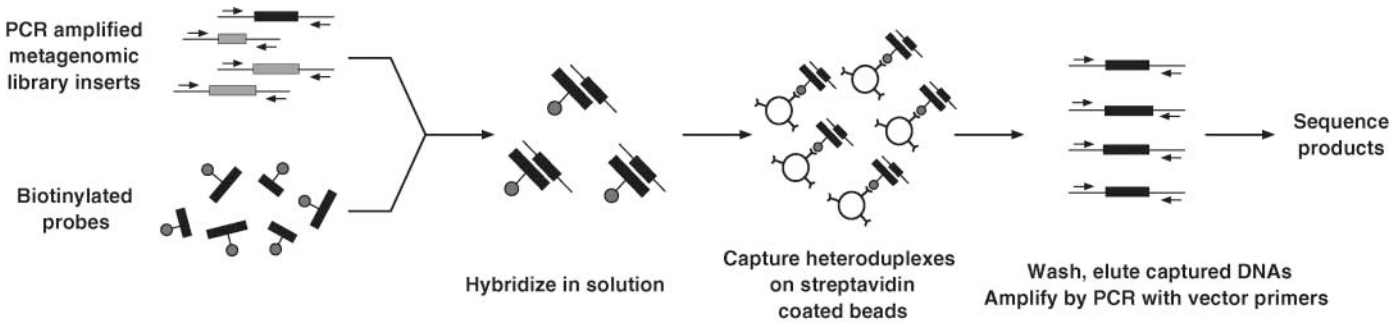

Fig. 7. Recovery of Neanderthal genomic sequences from library NE1 by direct genomic selection.

success in recovering both previously unknown cave bear and known Neanderthal genomic sequences using direct genomic selection indicates that this is a feasible strategy for purifying specific cloned Neanderthal sequences out of a high background of Neanderthal and contaminating microbial DNA. This raises the possibility that, should multiple Neanderthal metagenomic libraries be constructed from independent samples, direct selection could be used to recover Neanderthal sequences from several individuals to obtain and confirm important human-specific and Neanderthal-specific substitutions.

Conclusions. The current state of our knowledge concerning Neanderthals and their relationship to modern humans is largely inference and speculation based on archaeological data and a limited number of hominid remains. In this study, we have demonstrated that Neanderthal genomic sequences can be recovered using a metagenomic library-based approach and that specific Neanderthal sequences can be obtained from such libraries by direct selection. Our study thus provides a framework for the rapid recovery of Neanderthal sequences of interest from multiple independent specimens, without the need for whole-genome resequencing. Such a collection of targeted Neanderthal sequences would be of immense value for understanding human and Neanderthal biology and evolution. Future Neanderthal genomic studies, including targeted and whole-genome shotgun sequencing, will provide insight into the profound phenotypic divergence of humans both from the great apes and from our extinct hominid relatives, and will allow us to explore aspects of Neanderthal biology not evident from artifacts and fossils.

References and Notes

1. P. Mellars, *Nature* **432**, 461 (2004).
2. F. H. Smith, E. Trinkaus, P. B. Pettitt, I. Karavanic, M. Paunovic, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12281 (1999).
3. M. Krings *et al.*, *Cell* **90**, 19 (1997).
4. M. Krings *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5581 (1999).
5. S. Pääbo *et al.*, *Annu. Rev. Genet.* **38**, 645 (2004).
6. D. Serre *et al.*, *PLoS Biol.* **2**, e57 (2004).
7. M. Currat, L. Excoffier, *PLoS Biol.* **2**, e421 (2004).
8. S. G. Tringe *et al.*, *Science* **308**, 554 (2005).
9. S. G. Tringe, E. M. Rubin, *Nat. Rev. Genet.* **6**, 805 (2005).
10. J. P. Noonan *et al.*, *Science* **309**, 597 (2005).
11. M. Margulies *et al.*, *Nature* **437**, 376 (2005).
12. H. N. Poinar *et al.*, *Science* **311**, 392 (2006).
13. Materials and methods are available as supporting material on *Science* Online.
14. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997).
15. Chimpanzee Sequencing and Analysis Consortium, *Nature* **437**, 69 (2005).
16. M. Hofreiter *et al.*, *Nucleic Acids Res.* **29**, 4793 (2001).
17. S. Kumar, A. Filipski, V. Swarna, A. Walker, S. B. Hedges, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18842 (2005).
18. N. Patterson, D. Richter, S. Gnerre, E. Lander, D. Reich, *Nature*, in press; published online 17 May 2006 (10.1038/nature04789).
19. The International HapMap Consortium *et al.*, *Nature* **437**, 1299 (2005).
20. B. F. Voight *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18508 (2005).
21. A. M. Adams, R. R. Hudson, *Genetics* **168**, 1699 (2004).
22. I. McDougall *et al.*, *Nature* **433**, 733 (2005).
23. V. Plagnol, J. D. Wall, *PLoS Genet.*, in press (110.1371/journal.pgen.0020105.eor).
24. S. Bashiardes *et al.*, *Nat. Methods* **2**, 63 (2005).
25. P. Mellars, *Nature* **439**, 931 (2006).

26. Neanderthal sequences reported in this study have been deposited in GenBank under accession numbers DX935178 to DX936503. We thank E. Green, M. Lovett, and members of the Rubin, Pääbo, and Pritchard laboratories for insightful discussions and support. J.P.N. was supported by NIH National Research Service Award fellowship 1-F32-GM074367. G.C. and S.K. were supported by grant R01 HG002772-1 (NIH) to J.K.P. This work was supported by grant HL066681, NIH Programs for Genomic Applications, funded by the National Heart, Lung and Blood Institute; and by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract number DE-AC02-05CH11231.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5802/1113/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S12
References

16 June 2006; accepted 17 August 2006
10.1126/science.1131412

Downloaded from www.sciencemag.org on April 9, 2007

REPORTS

Resilient Machines Through Continuous Self-Modeling

Josh Bongard,[1]*† Victor Zykov,[1] Hod Lipson[1,2]

Animals sustain the ability to operate after injury by creating qualitatively different compensatory behaviors. Although such robustness would be desirable in engineered systems, most machines fail in the face of unexpected damage. We describe a robot that can recover from such change autonomously, through continuous self-modeling. A four-legged machine uses actuation-sensation relationships to indirectly infer its own structure, and it then uses this self-model to generate forward locomotion. When a leg part is removed, it adapts the self-models, leading to the generation of alternative gaits. This concept may help develop more robust machines and shed light on self-modeling in animals.

Robotic systems are of growing interest because of their many practical applications as well as their ability to help understand human and animal behavior (*1*–*3*), cognition (*4*–*6*), and physical performance (*7*). Although industrial robots have long been used for repetitive tasks in structured environments, one of the long-standing challenges is achieving robust performance under uncertainty (*8*). Most robotic systems use a manually constructed mathematical model that captures the robot's dynamics and is then used to plan actions (*9*). Although some parametric identification methods exist for automatically improving these models (*10*–*12*), making accurate models is difficult for complex machines, especially when trying to account for possible topological changes to the body, such as changes resulting from damage.

[1]Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853, USA. [2]Computing and Information Science, Cornell University, Ithaca, NY 14853, USA.

*Present address: Department of Computer Science, University of Vermont, Burlington, VT 05405, USA.
†To whom correspondence should be addressed. E-mail: josh.bongard@uvm.edu