



Neanderthal genomics and the evolution of modern humans

James P. Noonan

Genome Res. 2010 20: 547-553

Access the most recent version at doi:[10.1101/gr.076000.108](https://doi.org/10.1101/gr.076000.108)

References This article cites 60 articles, 21 of which can be accessed free at:
<http://genome.cshlp.org/content/20/5/547.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © 2010 by Cold Spring Harbor Laboratory Press

Neanderthal genomics and the evolution of modern humans

James P. Noonan¹

Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520, USA; Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut 06520, USA

Humans possess unique physical and cognitive characteristics relative to other primates. Comparative analyses of the human and chimpanzee genomes are beginning to reveal sequence changes on the human lineage that may have contributed to the evolution of human traits. However, these studies cannot identify the genetic differences that distinguish modern humans from archaic human species. Here, I will discuss efforts to obtain genomic sequence from Neanderthal, the closest known relative of modern humans. Recent studies in this nascent field have focused on developing methods to recover nuclear DNA from Neanderthal remains. The success of these early studies has inspired a Neanderthal genome project, which promises to produce a reference Neanderthal genome sequence in the near future. Technical issues, such as the level of Neanderthal sequence coverage that can realistically be obtained from a single specimen and the presence of modern human contaminating sequences, reduce the detection of authentic human–Neanderthal sequence differences but may be remedied by methodological improvements. More critical for the utility of a Neanderthal genome sequence is the evolutionary relationship of humans and Neanderthals. Current evidence suggests that the modern human and Neanderthal lineages diverged before the emergence of contemporary humans. A fraction of biologically relevant human–chimpanzee sequence differences are thus likely to have arisen and become fixed exclusively on the modern human lineage. A reconstructed Neanderthal genome sequence could be integrated into human–primate genome comparisons to help reveal the evolutionary genetic events that produced modern humans.

We are a young species. The most recent genetic and fossil evidence suggests that modern humans emerged in East Africa ~200,000 yr ago, a fraction of the estimated 6 million yr since the divergence of the human and chimpanzee lineages (Cavalli-Sforza and Feldman 2003; McDougall et al. 2005). In this brief time, modern humans have spread across the globe, far surpassing all other primate and archaic human species in sheer numbers, geographic range, technological sophistication, and impact on the environment. Our success is due in part to complex biological features, including increased brain size, bipedalism, and modifications in craniofacial and limb morphology, which began to emerge on the human lineage well before the rise of modern humans. These adaptations facilitated the evolution of uniquely human behavioral traits such as language and are ultimately due to DNA sequence changes arising on the human lineage since the human–chimpanzee split. Identifying these changes has become a major focus of human genetics and genomics.

The sequencing of the chimpanzee, rhesus macaque, and other primate genomes has made it possible to comprehensively identify human-specific sequence changes using comparative methods (Chimpanzee Sequencing and Analysis Consortium 2005; Rhesus Macaque Genome Sequencing and Analysis Consortium 2007). Recent studies have attempted to differentiate the critical functional changes from neutral events using various statistical methods to identify human genes and gene regulatory sequences showing unexpectedly rapid evolution compared to the inferred ancestral primate state (Clark et al. 2003; Chimpanzee Sequencing and Analysis Consortium 2005; Pollard et al. 2006a,b; Prabhakar et al. 2006, 2008; Haygood et al. 2007). The results from these analyses comprise a catalog of potential human-specific adaptive evolution

on the molecular level. However, these studies are unable to distinguish between sequence changes unique to modern humans and those shared between modern humans and earlier human species, and thus will not identify uniquely modern human genetic or biological features. Given our recent origin, the sequence changes that distinguish modern humans from archaic human species will likely be a small subset of all human–chimpanzee sequence differences.

These insights could in principle be obtained by examining genomic DNA directly from an archaic hominid that diverged from the modern human lineage prior to the emergence of contemporary humans. Neanderthal has been the primary target of such studies, for several reasons. Neanderthals are close relatives of anatomically modern humans, and both species are believed to have cohabited Europe for ~6000 yr, raising the possibility of interbreeding (Krings et al. 1997; Mellars 2004). Also, remains are available and are occasionally of a sufficient level of preservation that DNA can be recovered from them. Most efforts in this field have largely focused on PCR amplification of short mitochondrial sequences. However, several recent studies have employed a metagenomic approach coupled with massively parallel high-throughput sequencing to obtain genomic sequence from mammoth and Neanderthal remains (Green et al. 2006; Noonan et al. 2006; Poinar et al. 2006; Miller et al. 2008). Based on these early successes, a Neanderthal genome project was initiated in 2005. As of this writing, this project has produced a low coverage, draft genome sequence that has yet to be released (Dalton 2009). There has been considerable speculation about what this project will reveal about Neanderthals and their relationship to modern humans, as well as fanciful notions raised in the popular media of using a Neanderthal genome sequence to somehow resurrect the species. Here we will consider the technical and analytical issues involved and discuss what insights a Neanderthal genome sequence may provide about modern human evolution.

¹Corresponding author.

E-mail james.noonan@yale.edu.

Article is online at <http://www.genome.org/cgi/doi/10.1101/gr.076000.108>.

Rationale

The primary motivation behind generating a Neanderthal reference genome is to determine how distinct modern humans really are from all earlier versions of humanity. We are the only remaining human species, and thus we do not know if Neanderthals or our other extinct relatives shared our capacity for invention, abstract reasoning, or language. We have had to speculate on these matters based on the bones, the settlements, and the artifacts Neanderthals left behind. The question of modern human and Neanderthal biological similarity is particularly compelling given the recent common ancestry of both species: Based on both genomic and mitochondrial sequence comparisons, the lineages leading to modern humans and Neanderthals likely diverged in Africa ~300,000–700,000 yr ago (Krings et al. 1997; Serre et al. 2004; Green et al. 2006, 2008; Noonan et al. 2006). This genetic evidence has become folded into a narrative of modern human and Neanderthal evolutionary history that continues to frame comparative studies of both species. In its simplest form, the modern human and Neanderthal lineages continued on parallel evolutionary tracks subsequent to their divergence, with the descendants of one branch migrating to Europe and giving rise to Neanderthals, and the other branch remaining in Africa and eventually producing us (White et al. 2003; Mellars 2004; Hublin 2009; Tattersall 2009). The modern human colonization of Europe ~40,000 yr ago potentially brought both lineages back into widespread contact (Mellars 2004).

Given their very recent common ancestry, how much did the species have in common at this point? Were modern humans and Neanderthals capable of interbreeding, and, if so, did it happen to any appreciable extent? Or were the species so different that no meaningful exchange of information could occur? The first modern humans to colonize Europe had the same cognitive capacities as humans living today: They made cave paintings, figurines, and rudimentary musical instruments and most likely had language, indicating a highly developed modern human capacity for abstraction (Mellars 2004; Bailey et al. 2009; Conard 2009; Conard et al. 2009). Such symbolic behavior was not a new phenomenon; there is strong evidence that modern humans were making personal decorations and carving abstract representations in Africa at least ~75,000 yr ago (Henshilwood et al. 2002, 2004, 2009). We do not know if Neanderthals had similar talents; they certainly made tools and may have had larger brains than modern humans (Ponce de León et al. 2008). However, hypotheses that Neanderthals were independently capable of complex symbolic behavior, based on analyses of artifacts speculated to have been fashioned by Neanderthals, are disputed (Gravina et al. 2005; Zilhão et al. 2006; Mellars et al. 2007). We also do not know if Neanderthals had language, although arguments for and against have been made based on the significance of anatomical features shared between humans and Neanderthals that may or may not have enabled Neanderthals to speak (Houghton 1993; Lieberman 1994; Martinez et al. 2008). The fundamental problem in all of these studies is the lack of data: Modern humans came to dominate the planet, while Neanderthals died out ~28,000 yr ago (Finlayson et al. 2006). A reference Neanderthal genome sequence by itself will not resolve such questions. What it will provide, when compared with the modern human and chimpanzee reference genomes, is a list of genetic changes unique to modern humans or Neanderthals and changes relative to chimpanzee that are shared between both human species. This will facilitate the experimental discovery of biological differences between modern humans and

Neanderthals at a molecular level. The discovery of changes in gene expression or protein function unique to modern humans or Neanderthals will serve as an entry point for studying more complex biological phenomena in each lineage, including language and other behavioral traits.

Technical challenges, early studies, and solutions

Ancient DNA degradation and contamination

All studies of Neanderthal genomic DNA use material obtained from fossilized Neanderthal bones that are tens of thousands of years old. As one might expect, the quality of the material that can be recovered from such specimens is very poor, as DNA degrades over time. Consequently, there is wide variability in DNA preservation among available Neanderthal specimens, and most Neanderthal bones yield no usable DNA (Green et al. 2006). When present, Neanderthal genomic DNA is recovered in short (50- to 100-bp) fragments. The information content of Neanderthal DNA is also degraded: Individual nucleotides are subject to chemical modification, with deamination of cytosine to uracil the most common lesion (Pääbo et al. 2004; Briggs et al. 2007). Moreover, the fragments of Neanderthal genomic DNA are mixed with microbial contaminants that constitute >90% of the recovered DNA. The amount of DNA damage and microbial contamination are dependent on ambient environmental conditions: The ancient specimens that have provided the most intact DNA are mammoth remains recovered from permafrost. These specimens often include preserved hair shafts and soft tissues from which relatively high-quality DNA can be recovered (Poinar et al. 2006; Rogaev et al. 2006; Gilbert et al. 2007; Miller et al. 2008). None of the Neanderthal specimens providing the DNA for whole-genome sequencing approach this level of preservation (Green et al. 2006, 2008; Briggs et al. 2007). In addition to these challenges, ancient specimens frequently become contaminated with modern human DNA during handling and DNA extraction (Pääbo et al. 2004). This poses obvious problems for distinguishing modern human from Neanderthal DNA, since the frequency of single-base mismatches between the two genomes is estimated to be <0.5% (Noonan et al. 2006).

All ancient DNA studies must overcome these limitations. The earliest efforts to recover Neanderthal sequence relied on PCR amplification of mitochondrial DNA (Krings et al. 1997). Analysis of PCR-amplified mtDNA hypervariable region sequences from several Neanderthals suggested the modern human and Neanderthal lineages diverged over 500,000 yr ago, with little subsequent admixture (Krings et al. 1997; Serre et al. 2004). This was the first evidence that Neanderthals most likely did not interbreed with modern humans during their cohabitation of Europe. However, compared with genomic sequences, mtDNA provides limited insight into modern human and Neanderthal evolutionary history and no insight into Neanderthal biology. However, PCR was clearly not suitable for large-scale, high-throughput recovery of Neanderthal genomic DNA, so a new approach was required.

Metagenomic analysis of ancient DNA

The development of metagenomic approaches for characterizing complex DNA mixtures, in which high-throughput sequencing is used to comprehensively characterize all sequences present in a sample and the origins of each sequence are determined by comparison to known genomes, has made it possible to recover

significant amounts of Neanderthal genomic DNA (Tringali and Rubin 2005). The first study to apply a metagenomic strategy to ancient DNA used standard library construction methods to clone all DNA obtained from skeletal remains of two ~40,000-yr-old Pleistocene cave bears, followed by conventional Sanger sequencing massively parallelized on an industrial scale. Putative cave bear genomic sequences were identified based on comparison to the dog genome (Noonan et al. 2005). The results of this study highlighted the technical difficulties a Neanderthal genome project would face. Only 1.3% (26 kb out of 2.09 Mb) of all sequenced bases appeared to be cave bear genomic sequence. The remaining sequences were most similar to known prokaryotic genomes or other metagenomic samples such as soil or were from a source not represented in the sequence databases. However, this study suggested that brute force sequencing could be used to separate ancient genomic sequences from a species of interest from an extremely high background of microbial contaminants.

Short-read, massively parallel next-generation sequencing technologies generate enough raw sequence data to make whole genome-sequencing from ancient specimens technically feasible. The current standard in metagenomic analysis of ancient DNA is to eliminate the library construction step in favor of direct sequencing. The first such study used 454 Life Sciences (Roche) pyrosequencing to generate 13 Mb of mammoth genomic sequence from a well-preserved sample (Poinar et al. 2006). The same approach was subsequently used to generate a low coverage (0.7-fold) shotgun sequence of the mammoth genome (Miller et al. 2008). In both studies, a large fraction of the recovered DNA was mammoth (>90% in one case), due to the exceptional preservation of the samples. Subsequent ancient DNA studies have used much higher throughput sequencing platforms, such as the Illumina Genome Analyzer, that generate tens of millions of reads of up to 100 bp in a single run (Blow et al. 2008). These technologies are well suited for metagenomic analysis of highly fragmented ancient DNAs, and the Neanderthal genome sequence about to be released was generated using a combination of 454 and Illumina sequencing (Pennisi 2009).

The feasibility of Neanderthal genome sequencing was established by two complementary studies of DNA obtained from a ~38,000-yr-old Neanderthal specimen from Vindija Cave, Croatia (Green et al. 2006; Noonan et al. 2006). In one study, Neanderthal genomic DNA was cloned into libraries and sequenced using a combination of traditional Sanger sequencing and 454 pyrosequencing (Noonan et al. 2006). Approximately 1% of the reads obtained were putatively Neanderthal, yielding ~65 kb of sequence. When compared to human and chimpanzee, these reads exhibited a pattern of apparent Neanderthal-specific substitutions that were heavily biased toward C-to-T and G-to-A transitions, suggesting many of these events were errors induced by cytosine deamination. Analyses of human–Neanderthal divergence based on the putatively Neanderthal autosomal sequences obtained in this study indicated that the time to the most recent common ancestor (TMRCA) of the reference human genome and Vindija Neanderthal is ~700,000 yr and that the ancestral populations that gave rise to modern humans and Neanderthals split ~370,000 yr ago. No significant evidence of human–Neanderthal admixture was found. In the parallel study, DNA obtained from the same specimen was directly analyzed by massively parallel 454 pyrosequencing (Green et al. 2006). Approximately 6.2% (~1 Mb) of the sequences recovered appeared to be Neanderthal DNA, a somewhat greater yield than the related study obtained from the same sample. The estimated reference human genome–Vindija

Neanderthal TMRCA was also somewhat lower: ~516,000 yr, slightly larger than the divergence time the investigators calculated between two humans (~459,000 yr). At the time, this result suggested a considerable level of admixture between modern humans and Neanderthals. Subsequent analyses suggested that the putative Neanderthal DNA analyzed in this study was substantially contaminated with modern human DNA, an issue we will return to below (Wall and Kim 2007; Green et al. 2008). Despite such technical issues, these studies demonstrated that a Neanderthal genome sequence could be obtained.

Availability of samples

There is a finite supply of Neanderthal remains, and most of those are too poorly preserved to yield usable amounts of DNA (Green et al. 2006). The successful recovery of DNA from Neanderthal remains from central Asia and Siberia has increased the number of samples available for whole-genome sequencing and has provided insight into Neanderthal genetic diversity (Krause et al. 2007). However, it is unlikely that there will be more than a handful of low-coverage Neanderthal genome sequences for the foreseeable future. This will make it difficult to discriminate between Neanderthal-specific substitutions and common polymorphisms or to distinguish true cases of shared polymorphism between modern humans and Neanderthals from contamination with modern human DNA. Moreover, DNA extraction is a destructive process, and physical anthropologists are understandably reluctant to provide Neanderthal bones and teeth, which have great morphometric value, for ancient DNA studies. Consequently, several methods have been developed to determine whether a particular bone will yield DNA, to reduce the amount of material necessary for metagenomic sequencing, and to target specific sequences of interest (Pääbo et al. 2004; Noonan et al. 2006; Krause et al. 2007; Meyer et al. 2008; Briggs et al. 2009a,b). The decreasing cost and increasing throughput and sensitivity of sequencing technologies will eventually obviate the need for targeted sequence recovery. However, at present such methods are the best means to confirm the authenticity of specific sequences of interest identified in the handful of low-coverage Neanderthal genomes that will be available.

Authenticity

Ensuring that the putatively Neanderthal sequences obtained in metagenomic analyses represent the actual state of the Neanderthal genome is the most difficult remaining challenge in Neanderthal genomics. The two most important factors that compromise the authenticity of Neanderthal sequences are damage-induced mutation and contamination with modern human DNA. DNA damage irreversibly changes genetic information and, in comparisons of human, Neanderthal, and chimpanzee sequences, will result both in an overestimate of Neanderthal substitutions—as errors and substitutions will at first be indistinguishable—and in an underestimate of modern human-specific substitutions, in cases where the damage event changes the Neanderthal sequence to match the human state (Green et al. 2006; Noonan et al. 2006). The majority of damage-induced mutations are due to cytosine deamination and result in an excess of C-to-T and G-to-A transitions that can be readily detected in comparative analyses (Green et al. 2006; Noonan et al. 2006; Briggs et al. 2007). In principle, screening out errors from true Neanderthal-specific substitutions or polymorphisms can be readily accomplished by increasing

sequence coverage, as deamination events are unlikely to occur at high frequency at the same exact site in multiple independent ancient DNA fragments. To date this has been accomplished largely by cloning and sequencing multiple PCR products (Krings et al. 1997; Krause et al. 2007; Lalueza-Fox et al. 2007). Initial drafts of the reference Neanderthal genome sequence will be low coverage, so individual sites are likely to be represented by only one read. However, Neanderthal-specific substitutions or polymorphisms identified in draft genomic sequence can be confirmed by replication in independent samples, either by PCR or by whole-genome or targeted sequencing. Sequence coverage of single Neanderthal genomes is also likely to increase as sequencing throughput increases.

Mitigating the effect of contamination with modern human DNA is more difficult. The modern human and Neanderthal genomes are virtually identical (Green et al. 2006; Noonan et al. 2006). Most Neanderthal sequence fragments will therefore perfectly match the reference human genome and, thus, in principle, could be derived from either species. Moreover, modern human DNA can be amplified and sequenced much more efficiently than ancient Neanderthal DNA, so even small amounts of contamination can cause most of the hominid sequences obtained to be from modern human. Modern human DNA can be introduced into a Neanderthal specimen at several points: during the initial recovery of the bone, during subsequent handling, during DNA extraction, or during preparation of the sequencing library. Most available Neanderthal bones were collected without regard to their potential use as a source of DNA and have passed through many hands that have left their DNA behind. Rigorous procedures have been developed to prevent contamination during DNA extraction and to determine the level of contamination present in a particular specimen (Pääbo et al. 2004; Green et al. 2006). Neanderthal specimens are also being collected in the field using methods to prevent modern human DNA from being introduced (Krause et al. 2007). Despite these strategies, contamination can still be introduced during subsequent handling of the extracted DNA. This appears to have occurred in one of the pilot Neanderthal genome sequencing studies described above (Green et al. 2006). It is now accepted that a substantial amount (at least 11%) of the hominid DNA recovered in this study was derived from a modern human source, which resulted in an underestimate of the TMRCA of the reference human genome and Vindija Neanderthal sequence and in an overestimate of the degree of admixture (Wall and Kim 2007; Green et al. 2008). Since PCR amplification and sequencing of mtDNA indicated a very low level of modern human mitochondrial DNA in the initial extract, it was suspected that the contamination was introduced during construction of the sequencing library (Green et al. 2008, 2009). These observations have spurred the development of several quality control steps in Neanderthal DNA handling and sequencing, including ligating adapters to Neanderthal DNA immediately after extraction to provide molecular barcodes that will distinguish it from modern DNA potentially introduced in downstream applications (Green et al. 2009).

Nevertheless, it will be impossible to completely exclude the possibility of contamination. The primary consequence of contamination on human–Neanderthal sequence comparisons will be to underestimate the number of human-specific substitutions. Metrics to evaluate the level of contamination in Neanderthal genome sequence—and thus the probability that a particular perfect sequence match between human and Neanderthal is authentic—include measuring the frequency of human-specific mtDNA and Y-chromosome sequences (Green et al. 2008, 2009). This

strategy assumes that the level of mitochondrial or Y-chromosome contamination is similar to the average level of autosomal DNA contamination. Rigorous statistical methods to model human contamination in Neanderthal samples have also been proposed (Coop et al. 2008). However, the gold standard for ruling out modern human contamination is the same as for damage-induced mutations: replication, either by increased sequence coverage or by sequencing multiple Neanderthal specimens. Until that is accomplished, all instances of human–Neanderthal sequence identity suggested by the draft Neanderthal genome sequence must be considered preliminary.

Comparative analysis of human and Neanderthal genomes

Maps of human and Neanderthal sequence differences

The draft Neanderthal genome sequence consists largely of 35- to 100-bp sequence reads that provide $\sim 1\times$ coverage (Pennisi 2009). De novo sequence assembly is not possible with such data—in fact, it is a daunting computational challenge even with $>40\times$ coverage of a modern human genome (Simpson et al. 2009). Initial human, Neanderthal, and chimpanzee sequence comparisons will thus involve tiling Neanderthal reads onto the human and chimpanzee reference genomes. These analyses will largely be limited to detecting lineage-specific substitutions and small insertions or deletions (indels). Some insight may be gained into segmental duplications or other large-scale genome rearrangements, but it is not clear how this can be achieved. The draft Neanderthal genome will thus provide three novel data sets: a genome-wide map of substitutions and indels that appear to be specific to the modern human reference genome relative to Neanderthal and chimpanzee; a map of Neanderthal-specific changes, which will be dominated by damage-induced errors; and a map of shared human–Neanderthal-derived changes relative to chimpanzee, which may include modern human contaminants (Fig. 1A). The most informative of these will be the map of apparent modern human-specific substitutions, which will complement existing maps of human-specific fixed sequence differences obtained from human–chimpanzee comparisons. Coupled with existing data sets of human polymorphisms and the large number of individual human genomes currently being sequenced, these maps will identify sequence differences arising since the modern human–Neanderthal divergence that are fixed in all humans.

Analyses of human and Neanderthal polymorphism are somewhat more complicated. Many human polymorphisms are expected to predate the human–Neanderthal split and thus may be present in Neanderthal as well. Consequently, fixed substitutions in modern humans may be polymorphic in Neanderthal, as may be sites where human and Neanderthal are derived relative to chimpanzee. Human polymorphisms where the Neanderthal reference genome has the ancestral allele may nevertheless be polymorphic in Neanderthal; this will confound efforts to date human polymorphisms based on the Neanderthal allele state. It is presently not clear how the effects of Neanderthal polymorphism on substitution maps will be addressed. Genome-wide insight into Neanderthal genetic diversity will not be available for some time and, even then, will be limited to a few individuals.

Sequences of interest

Most modern human-specific sequence changes will be functionally neutral. The challenge will be to identify the subset of changes

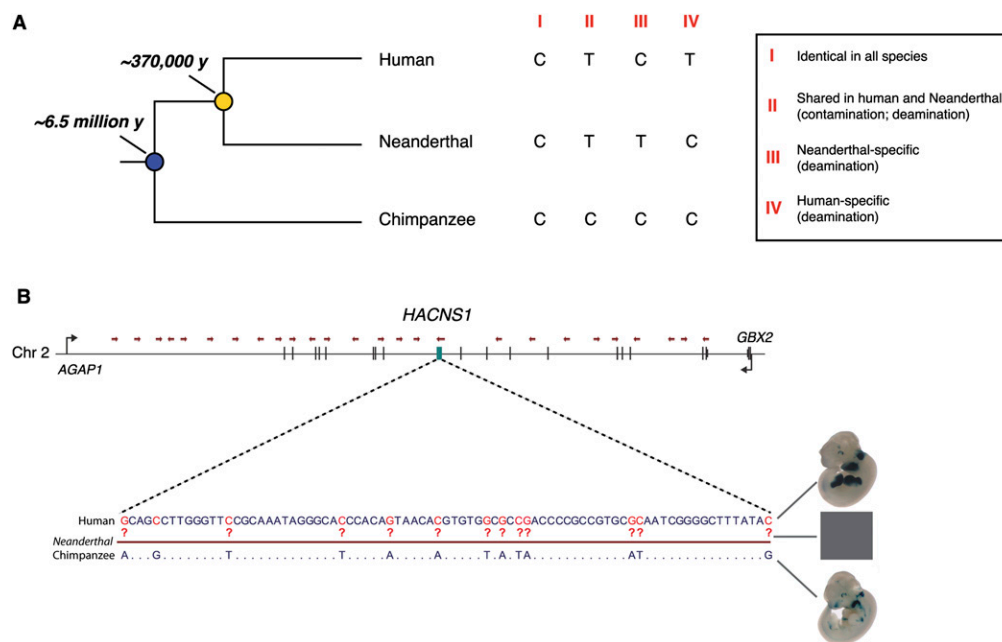


Figure 1. (A) Evolutionary relationship of modern human, Neanderthal, and chimpanzee. Divergence dates shown are estimated from genome sequence comparisons. A schematic of the single nucleotide changes detected in human–Neanderthal–chimpanzee comparisons is shown at *right*, ordered by the frequency at which they are likely to appear. Substitution classes and the ancient DNA artifacts particularly problematic for each are described in the legend. (B) Value of Neanderthal genome sequence for dating functionally relevant human-specific substitutions, using the *HACNS1* enhancer as an example. (*Top*) *HACNS1* is located in an intron of *AGAP1* and downstream of *GBX2* on human chromosome 2. (*Bottom*) The 13 human-specific substitutions implicated in the human-specific gain of function in this element. The enhancer expression patterns conferred by *HACNS1* and the chimpanzee ortholog in E11.5 mouse embryos are shown on the *right*. In a human–Neanderthal genome comparison, single Neanderthal sequence reads (red arrows) will be tiled across this region. A single read spanning *HACNS1* may identify which substitutions are unique to modern humans and which are shared between modern humans and Neanderthals. This analysis would also provide insight into the enhancer expression pattern of the Neanderthal *HACNS1* ortholog.

that are relevant to modern human evolution. In the past decade, genome-wide efforts to identify sequences under accelerated evolution on the human lineage based on substitution rates, or that show evidence of recent positive selection based on patterns of polymorphism, have provided a long list of sequences that may encode human-specific functions. Among these are genes such as *FOXP2*, which contains two human-specific amino acid substitutions that were initially attributed to an episode of positive selection early in modern human evolution (~200,000 yr ago) (Enard et al. 2002). *FOXP2* is of particular interest as it is suspected of having a role in the evolution of language, which is widely considered to be unique to modern humans (Lai et al. 2001; Enard et al. 2002). The timing of the selective episode at *FOXP2* has been called into question based on the finding that the human-specific amino acid changes are apparently present in Neanderthals, although this result has also been disputed (Krause et al. 2007; Coop et al. 2008). The validity of the *FOXP2* Neanderthal result and the overall relevance of *FOXP2* to human evolution are beyond the scope of this Perspective, but these serve as illustrations of how Neanderthal genome sequence can be used. Another example is the melanocortin 1 receptor (*MC1R*), inactivating mutations which contribute to pale skin and red hair in humans; a Neanderthal-specific variant in this gene was recently identified that reduces the activity of the resulting protein, suggesting that Neanderthals may have varied in pigmentation (Lalueza-Fox et al. 2007). All genes with human-specific amino acid changes are potentially interesting targets for human–Neanderthal comparative analysis (Chimpanzee Sequencing and Analysis Consortium 2005). What may be an unremarkable number of human-specific changes relative to chimpanzee, given the evolutionary distance of those species, may turn

out to be modern human-specific accelerated evolution relative to Neanderthal. Neanderthal genome sequence would also provide insight into the origin of human-specific unprocessed pseudogenes, which may include adaptive instances of gene loss in human evolution (Wang et al. 2006). Finally, Neanderthal genome sequence may help detect episodes of recent positive selection in the genome, through the identification of unusually long runs of linked human-specific derived substitutions that may be due to selective sweeps occurring early in modern human evolution. This approach would augment current efforts that use modern human polymorphism data to identify selective sweeps, and by providing a map of human-specific substitutions in such regions would help pinpoint the functional substitutions fixed by selection (Voight et al. 2006; Sabeti et al. 2007). However, as described above our limited understanding of Neanderthal polymorphism may complicate such analyses.

A Neanderthal reference genome would also complement ongoing efforts to annotate uniquely human *cis*-regulatory functions in the genome. Several recent studies have identified known or putative regulatory sequences that show high rates of nucleotide substitution on the human lineage, which may be due to positive selection for changes in regulatory functions (Pollard et al. 2006a,b; Prabhakar et al. 2006; Haygood et al. 2007). These efforts have largely focused on promoters and highly conserved noncoding sequences (CNSs), as CNSs are known to be enriched in transcriptional enhancers with *in vivo* developmental activities (Pennacchio et al. 2006; Visel et al. 2008). For example, a recent global study of human promoters found evidence of positive selection in promoters of genes involved in neural development and nutrition (Haygood et al. 2007). Another study surveyed patterns of human-specific nucleotide substitution in >110,000 CNSs in the genome and

identified 992 that showed statistically significant evidence of evolutionary acceleration (Prabhakar et al. 2006). These human-accelerated conserved noncoding sequences (HACNSs), which are known to include developmental enhancers, were overrepresented near genes involved in neuronal cell adhesion. This finding, along with the observation that the promoters of neuronal genes show evidence of human-specific positive selection, suggests that *cis*-regulatory changes may have had a profound impact on human brain evolution. Neanderthal genome sequence would also provide insight into the evolutionary origin of regulatory RNAs with putative human-specific functions: In one study of human-specific sequence acceleration in transcribed and non-transcribed sequences that identified 49 human accelerated regions (HARs), the most accelerated element, *HAR1*, is part of a noncoding RNA expressed in the developing human brain (Pollard et al. 2006b).

Comparison of the human and Neanderthal reference genomes would also place bounds on the evolutionary age of individual sequence changes that generate human-specific functions. For example, one of the most rapidly evolving CNSs in the genome, identified as *HACNS1* or *HAR2* in the studies described above, has accumulated 16 human-specific substitutions despite being highly conserved in all nonhuman terrestrial vertebrates (Fig. 1B; Prabhakar et al. 2008). *HACNS1* functions as a transcriptional enhancer in multiple structures in embryonic day (E) 11.5 and E13.5 mouse embryos, including in the developing anterior limb (Prabhakar et al. 2008). However, the orthologous chimpanzee and rhesus enhancers fail to drive expression in the limb, suggesting *HACNS1* has gained human-specific function *in vivo*. Notably, humanized chimpanzee enhancers, in which 13 of the 16 human-specific substitutions were introduced into the chimpanzee sequence, drove expression in the limb, indicating that the substitutions identified in the computational screen were directly responsible for the functional change. These substitutions could have arisen at any point since the divergence of the human and chimpanzee lineages. Comparing *HACNS1* with its Neanderthal ortholog would establish whether some or all of these substitutions—and the functions they confer—are unique to modern humans (Fig. 1B).

These studies illustrate a general strategy toward an understanding of biological differences between modern humans and Neanderthals, in which the first step is the reverse genetic analysis of genes and gene regulatory elements showing human-specific or Neanderthal-specific sequence changes. In this approach, changes in basic molecular functions, such as enhancer activity, protein-DNA interactions, or receptor-ligand binding affinity are identified in synthetic assays. The phenotypic consequences of these molecular changes can then be assessed in mouse models: A recent study describing the introduction of a “humanized” version of *FOXP2* into the mouse genome by gene targeting is one early example (Enard et al. 2009). The data from such studies, combined with a growing body of information on human gene function, the effects of genetic variation on human phenotypes, and comprehensive efforts to functionally annotate the human genome, would provide the foundation for more sophisticated hypotheses concerning the biological similarity of modern humans and Neanderthals than can be generated from the paleoanthropological record alone.

Conclusion: How “human” were the Neanderthals?

Despite many technical and analytical challenges, a Neanderthal genome sequence has been generated and is certain to advance our understanding of modern human evolution. However, the prospect of a Neanderthal genome sequence has also spurred considerable

speculation about whether it would be possible to resurrect the species. Suggestions on how this might be accomplished include making the chimpanzee genome more Neanderthal-like (Wade 2009). The technical feasibility of such an experiment is dubious, and the number of chimpanzee embryos and foster mothers that would in principle be required make it ethically untenable. Statements that the Neanderthal genome sequence should be used to clone the species therefore undermine the credibility of the entire enterprise. Moreover, these speculations are based on the notion that Neanderthals were in some sense subhuman and that the moratoriums on human reproductive cloning should therefore not apply to them. Thus we return to the question at the heart of the Neanderthal genome project: Are modern humans fundamentally different from all archaic human species? Neanderthals were certainly intelligent: They fashioned recognizably human tools and lived in established communities, which alone suggests they were sufficiently “human” that we should not consider cloning them. It is hoped that direct comparison of the human and Neanderthal genomes will reveal whether our talents for invention, language, and abstract thought—the basis of our perceived uniqueness among species—are truly unique to us. It may be that Neanderthals lacked essential modern human characteristics. However, there is a long road ahead, and we should not be surprised if we discover that we are not as different from our extinct relatives as we believe.

Acknowledgments

I thank members of the Noonan laboratory for insightful discussions, and three anonymous reviewers for helpful suggestions on the manuscript. This work was supported in part by funds from the Edward J. Mallinckrodt foundation, the Kavli Institute for Neuroscience, and the Yale University School of Medicine. The author is not involved with the ongoing Neanderthal genome project and declares no conflicts of interest.

References

- Bailey SE, Weaver TD, Hublin JJ. 2009. Who made the Aurignacian and other early Upper Paleolithic industries? *J Hum Evol* **57**: 11–26.
- Blow MJ, Zhang T, Woyke T, Speller CF, Krivoshapkin A, Yang DY, Derevianko A, Rubin EM. 2008. Identification of ancient remains through genomic sequencing. *Genome Res* **18**: 1347–1353.
- Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, et al. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci* **104**: 14616–14621.
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z, et al. 2009a. Targeted retrieval and analysis of five Neanderthal mtDNA genomes. *Science* **325**: 318–321.
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Paabo S. 2009b. Primer extension capture: Targeted sequence retrieval from heavily degraded DNA sources. *J Vis Exp* 1573.
- Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat Genet* **33** (Suppl): 266–275.
- Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.
- Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, et al. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**: 1960–1963.
- Conard NJ. 2009. A female figurine from the basal Aurignacian of Hohle Fels Cave in southwestern Germany. *Nature* **459**: 248–252.
- Conard NJ, Malina M, Münzel SC. 2009. New flutes document the earliest musical tradition in southwestern Germany. *Nature* **460**: 737–740.
- Coop G, Bullaughey K, Luca F, Przeworski M. 2008. The timing of selection at the human *FOXP2* gene. *Mol Biol Evol* **25**: 1257–1259.
- Dalton R. 2009. Neanderthal genome to be unveiled. *Nature* **457**: 645.

- Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S. 2002. Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**: 869–872.
- Enard W, Gehre S, Hammerschmidt K, Holter SM, Blass T, Somel M, Bruckner MK, Schreiweis C, Winter C, Sohr R, et al. 2009. A humanized version of *Foxp2* affects cortico-basal ganglia circuits in mice. *Cell* **137**: 961–971.
- Finlayson C, Pacheco FG, Rodriguez-Vidal J, Fa DA, Gutierrez Lopez JM, Santiago Perez A, Finlayson G, Allue E, Baena Preysler J, Caceres I, et al. 2006. Late survival of Neanderthals at the southernmost extreme of Europe. *Nature* **443**: 850–853.
- Gilbert MT, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, Tikhonov A, Dalen L, Kuznetsova T, Kosintsev P, et al. 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317**: 1927–1930.
- Gravina B, Mellars P, Ramsey CB. 2005. Radiocarbon dating of interstratified Neanderthal and early modern human occupations at the Chatelperronian type-site. *Nature* **438**: 51–56.
- Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, et al. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**: 330–336.
- Green RE, Malaspina AS, Krause J, Briggs AW, Johnson PL, Uhler C, Meyer M, Good JM, Maricic T, Stenzel U, et al. 2008. A complete Neanderthal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**: 416–426.
- Green RE, Briggs AW, Krause J, Prufer K, Burbano HA, Siebauer M, Lachmann M, Paabo S. 2009. The Neanderthal genome and ancient DNA authenticity. *EMBO J* **28**: 2494–2502.
- Haygood R, Fedrigo O, Hanson B, Yokoyama K, Wray G. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet* **39**: 1140–1144.
- Henshilwood CS, d'Errico F, Yates R, Jacobs Z, Tribolo C, Duller GA, Mercier N, Sealy JC, Valladas H, Watts I, et al. 2002. Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science* **295**: 1278–1280.
- Henshilwood C, d'Errico F, Vanhaeren M, van Niekerk K, Jacobs Z. 2004. Middle Stone Age shell beads from South Africa. *Science* **304**: 404.
- Henshilwood CS, d'Errico F, Watts I. 2009. Engraved ochres from the Middle Stone Age levels at Blombos Cave, South Africa. *J Hum Evol* **57**: 27–47.
- Houghton P. 1993. Neanderthal supralaryngeal vocal tract. *Am J Phys Anthropol* **90**: 139–146.
- Hublin JJ. 2009. Out of Africa: Modern human origins special feature: The origin of Neandertals. *Proc Natl Acad Sci* **106**: 16022–16027.
- Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, Burbano HA, Hublin JJ, Hanni C, Forster J, de la Rasilla M, et al. 2007. The derived *FOXP2* variant of modern humans was shared with Neandertals. *Curr Biol* **17**: 1908–1912.
- Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S. 1997. Neanderthal DNA sequences and the origin of modern humans. *Cell* **90**: 19–30.
- Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**: 519–523.
- Lalueza-Fox C, Rompler H, Caramelli D, Staubert C, Catalano G, Hughes D, Rohland N, Pili E, Longo L, Condemi S, et al. 2007. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* **318**: 1453–1455.
- Lieberman P. 1994. Functional tongues and neanderthal vocal tract reconstruction: A reply to Dr. Houghton (1993). *Am J Phys Anthropol* **95**: 443–450.
- Martinez I, Arsuaga JL, Quam R, Carretero JM, Gracia A, Rodriguez L. 2008. Human hyoid bones from the middle Pleistocene site of the Sima de los Huesos (Sierra de Atapuerca, Spain). *J Hum Evol* **54**: 118–124.
- McDougall I, Brown FH, Fleagle JG. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* **433**: 733–736.
- Mellars P. 2004. Neanderthals and the modern human colonization of Europe. *Nature* **432**: 461–465.
- Mellars P, Gravina B, Bronk Ramsey C. 2007. Confirmation of Neanderthal/modern human interstratification at the Chatelperronian type-site. *Proc Natl Acad Sci* **104**: 3657–3662.
- Meyer M, Briggs AW, Maricic T, Hober B, Hoffner B, Krause J, Weihmann A, Pääbo S, Hofreiter M. 2008. From micrograms to picograms: Quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res* **36**: e5. doi: 10.1093/nar/gkm1095.
- Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard MD, Zhao F, Sher A, et al. 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**: 387–390.
- Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Dettler JC, Pääbo S, Rubin EM. 2005. Genomic sequencing of Pleistocene cave bears. *Science* **309**: 597–599.
- Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Pääbo S, Pritchard JK, et al. 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**: 1113–1118.
- Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. 2004. Genetic analyses from ancient DNA. *Annu Rev Genet* **38**: 645–679.
- Pennacchio L, Ahituv N, Moses A, Prabhakar S, Nobrega M, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis K, et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**: 499–502.
- Pennisi E. 2009. Neanderthal genomics. Tales of a prehistoric human genome. *Science* **323**: 866–871.
- Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, et al. 2006. Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science* **311**: 392–394.
- Pollard K, Salama S, King B, Kern A, Dreszer T, Katzman S, Siepel A, Pedersen J, Bejerano G, Baertsch R, et al. 2006a. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* **2**: e168. doi: 10.1371/journal.pgen.0020168.
- Pollard K, Salama S, Lambert N, Lambot M, Coppens S, Pedersen J, Katzman S, King B, Onodera C, Siepel A, et al. 2006b. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**: 167–172.
- Ponce de León MS, Golovanova L, Doronichev V, Romanova G, Akazawa T, Kondo O, Ishida H, Zollikofer CP. 2008. Neanderthal brain size at birth provides insights into the evolution of human life history. *Proc Natl Acad Sci* **105**: 13764–13768.
- Prabhakar S, Noonan JP, Pääbo S, Rubin EM. 2006. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**: 786.
- Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Morrison H, Fitzpatrick DR, Afzal V, et al. 2008. Human-specific gain of function in a developmental enhancer. *Science* **321**: 1346–1350.
- Rhesus Macaque Genome Sequencing and Analysis Consortium. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**: 222–234.
- Rogaev EI, Moliaka YK, Malyarchuk BA, Kondrashov FA, Derenko MV, Chumakov I, Grigorenko AP. 2006. Complete mitochondrial genome and phylogeny of Pleistocene mammoth *Mammuthus primigenius*. *PLoS Biol* **4**: e73. doi: 10.1371/journal.pbio.0040073.
- Sabeti P, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne E, McCarroll S, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**: 913–918.
- Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Mennecier P, Hofreiter M, Pospisil G, Pääbo S. 2004. No evidence of Neanderthal mtDNA contribution to early modern humans. *PLoS Biol* **2**: e57. doi: 10.1371/journal.pbio.0020057.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: A parallel assembler for short read sequence data. *Genome Res* **19**: 1117–1123.
- Tattersall I. 2009. Out of Africa: Modern human origins special feature: Human origins: Out of Africa. *Proc Natl Acad Sci* **106**: 16018–16021.
- Tringali SG, Rubin EM. 2005. Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* **6**: 805–814.
- Visel A, Prabhakar S, Akiyama J, Shoukry M, Lewis K, Holt A, Plajzer-Frick I, Afzal V, Rubin E, Pennacchio L. 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet* **40**: 158–160.
- Voight B, Kudaravalli S, Wen X, Pritchard J. 2006. A map of recent positive selection in the human genome. *PLoS Biol* **4**: e72. doi: 10.1371/journal.pbio.0040072.
- Wade N. 2009. Scientists in Germany draft Neanderthal genome. *The New York Times* February 12, A12.
- Wall JD, Kim SK. 2007. Inconsistencies in Neanderthal genomic DNA sequences. *PLoS Genet* **3**: 1862–1866.
- Wang X, Grus WE, Zhang J. 2006. Gene losses during human origins. *PLoS Biol* **4**: e52. doi: 10.1371/journal.pbio.0040052.
- White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC. 2003. Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* **423**: 742–747.
- Zilhão J, d'Errico F, Bordes JG, Lenoble A, Texier JP, Rigaud JP. 2006. Analysis of Aurignacian interstratification at the Chatelperronian-type site and implications for the behavioral modernity of Neandertals. *Proc Natl Acad Sci* **103**: 12643–12648.