Genomic Comparisons of Humans and Chimpanzees

Ajit Varki[1] and David L. Nelson[2]

[1]Glycobiology Research and Training Center, Departments of Medicine and Cellular and Molecular Medicine, University of California at San Diego, La Jolla, California 92093-0687; email: a1varki@ucsd.edu

[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030; email: nelson@bcm.tmc.edu

Annu. Rev. Anthropol. 2007. 36:191–209

First published online as a Review in Advance on June 18, 2007

The *Annual Review of Anthropology* is online at anthro.annualreviews.org

This article's doi:
10.1146/annurev.anthro.36.081406.094339

0084-6570/07/1021-0191$20.00

Key Words

human origins, anthropogeny, genetics, evolution, hominid

Abstract

The genome consists of the entire DNA present in the nucleus of the fertilized embryo, which is then duplicated in every cell in the body. A draft sequence of the chimpanzee genome is now available, providing opportunities to better understand genetic contributions to human evolution, development, and disease. Sequence differences from the human genome were confirmed to be ~1% in areas that can be precisely aligned, representing ~35 million single base-pair differences. Some 45 million nucleotides of insertions and deletions unique to each lineage were also discovered, making the actual difference between the two genomes ~4%. We discuss the opportunities and challenges that arise from this information and the need for comparison with additional species, as well as population genetic studies. Finally, we present a few examples of interesting findings resulting from genome-wide analyses, candidate gene studies, and combined approaches, emphasizing the pros and cons of each approach.

INTRODUCTION

We provide an overview of recent knowledge regarding genomic differences between humans and chimpanzees and focus on some questions and opportunities arising. To ensure accessibility to a broad range of readers, the emphasis is on general concepts rather than technical details. The interested reader is encouraged to consult the literature cited and the sources in the sidebar Online Sources of Information Regarding the Chimpanzee Genome. At the outset, we also emphasize that the sequencing of a genome does not automatically provide any biologically interesting answers. Rather, the importance of genome sequencing for biology is similar to deciphering the Periodic Table of Elements for chemistry or finding the Rosetta Stone for linguistics: It provides only the much-needed resource with which one can begin to ask questions of interest in fields such as anthropology. Thus, although absolutely necessary as a starting point, human and chimpanzee genomic sequencing are only the beginning of a long road of exciting explorations.

ONLINE SOURCES OF INFORMATION REGARDING THE CHIMPANZEE GENOME

1. Special issue of *Nature* on the chimpanzee genome: **http://www.nature.com/nature/focus/chimpgenome/index.html**
2. Chimpanzee Genome Resources at NCBI, NLM: **http://www.ncbi.nlm.nih.gov/genome/guide/chimp/**
3. An overview of chimpanzee genome sequencing at NHGRI (includes the original white paper proposals for the sequencing): **http://www.genome.gov/11008056**
4. *Pan troglodytes* Genome Browser Gateway at UC Santa Cruz: **http://genome.ucsc.edu/cgi-bin/hgGateway?org=Chimp&db=0**
5. Ensembl Chimpanzee Genome Server: **http://www.ensembl.org/Pan_troglodytes/index.html**
6. The Human Genome Sequencing Center at Washington University School of Medicine: **http://genome.wustl.edu/genome.cgi?GENOME=Pan%20troglodytes**
7. The Broad Institute at Massachusetts Institute of Technology and Harvard University: **http://www.broad.mit.edu/mammals/chimp/**
8. The Human Genome Sequencing Center at Baylor College of Medicine: **http://www.hgsc.bcm.tmc.edu/projects/chimpanzee/**

EXPLAINING HUMAN UNIQUENESS

Anthropogeny—*the study of the generation of man* (*Hooper Med. Dict.* 1839) or *the investigation of the origin of humans* (*Oxford Engl. Dict.* 2006)—is one of our greatest scientific challenges and is a subject of interest to most humans. Given the final goal, it seems reasonable to approach the problem in a human-focused manner, using information from all relevant areas of inquiry (**Figure 1**). First, from where did we originate? Our closest living evolutionary relatives are the common chimpanzee (*Pan troglodytes*) and the bonobo (*Pan paniscus*) (Goodman 1999). Combined molecular and fossil data suggest that the last common ancestor of humans within the chimpanzee-bonobo clade lived ~6–7 Mya (Chen & Li 2001, Brunet et al. 2002). Our next closest living relatives are the gorilla (*Gorilla gorilla*, common ancestor ~8 Mya) and the orangutan (*Pongo pygmaeus*, common ancestor ~13 Mya). Comparisons between humans and these "great ape" hominids are useful in understanding human origins (Goodman 1999, Gagneux & Varki 2001, Klein & Takahata 2002, Carroll 2003, Olson & Varki 2003, Enard & Paabo 2004, Gagneux 2004, Ruvolo 2004, Goodman et al. 2005, Li & Saunders 2005, McConkey & Varki 2005, Varki & Altheide 2005). The term great apes (including chimpanzees, bonobos, gorillas, and orangutans) is used here in the now colloquial sense because phylogenetic analysis of genomic information no longer supports this classical species grouping (Goodman 1999)—under the currently common classification, these species are now grouped together with humans in the family Hominidae.

The fossil record, consisting predominantly of fossils of human ancestors and

related species, has provided extensive data regarding the evolution of human ancestors (Wood & Collard 1999); however, investigators have found only one known example of a fossil in the chimpanzee lineage (McBrearty & Jablonski 2005). Additional clues arise from studies of the effects of environment and culture on the development and behavior of humans and other living hominids and comparative studies of the interactions of members of each species, particularly interactions between the two sexes, and between adults and developing progeny (**Figure 1**). This review provides an update on one specific approach to anthropogeny: comparing the genomes of humans and chimpanzees.

WHY THE CHIMPANZEE GENOME?

The genome is the sum total of double-stranded DNA present in the nucleus of the fertilized embryo, and which is subsequently duplicated in every other cell in the body. Mammalian nuclear genomes are roughly three billion nucleotides ("letters") in length, duplicated into pairs of chromosomes. Whereas the human genome is divided into 23 pairs of chromosomes, the chimpanzee and other great apes carry 24 pairs. Along with the smaller amount of DNA in mitochondria, these sequences represent the genetic instructions to generate a human or a chimpanzee—of course, in complex interactions with the physical, biological, conspecific, and cultural environments. The past two decades have witnessed remarkable progress in sequencing and understanding the human genome and its variation among individuals. The fraction of the human genome that is amenable to sequence analysis was found to be 2.85 billion nucleotides in length, and millions of single nucleotide polymorphisms (nucleotides found to vary from one individual to another with a population frequency of 1% or higher) have been defined (Int. Hum. Genome Seq. Consort. 2004, Consortium 2005). Global variation between humans at the single nucleotide level has been estimated at ~0.1%. The human genome carries a variety of repeated sequences, classified as short and long interspersed repeats, transposable elements (so-called jumping genes) such as retrovirus remnants, and satellite sequences (short, tandemly repeated sequences). Together these comprise more than 40% of the genome. Some 20–25,000 protein-encoding genes were found, but the coding regions of these comprise only ~1% of the genome.

A logical follow-up was to sequence genomes of other related animals to better understand evolutionary biology and unique features of humans (Thomas et al. 2003). The next genomes sequenced were those of mice and rats, experimental representatives of Rodentia, which also happens to be the closest evolutionary order within mammals to primates. The next phase proposed sequencing of primate genomes, and initial controversy developed regarding which genome would be most useful. Some favored an Old World monkey (VandeBerg et al. 2000) because such species had already been widely studied as biomedical models, and others favored the chimpanzee (McConkey et al. 2000, McConkey & Varki 2000b, Varki 2000). In reality, no single primate genome is sufficient to interpret the significance of human genomic differences (McConkey & Varki 2000a), and sequencing of multiple primate genomes is currently underway. This review focuses on what we have learned from the draft sequence of the chimpanzee genome (Chimpanzee Seq. Anal. Consort. 2005).

Single nucleotide polymorphisms: nucleotides found to vary from one individual to another with a population frequency of 1% or higher

WHAT CAN GENOMIC COMPARISONS TELL US?

The essentially complete human reference genome (Int. Hum. Genome Seq. Consort. 2004), coupled with the draft chimpanzee genome, allowed direct comparison of large fractions of these two data sets for interesting changes. Sequence differences were confirmed to be ~1% in regions that could be precisely aligned (lined up against each other,

indicating a common ancestral origin). This percentage represented ~35 million single base-pair differences. A surprising finding was that an additional ~45 million nucleotides of insertions and deletions were unique to each species, making the actual difference between the two genomes ~4%. Britten (2002) actually predicted this earlier.

Defining differences in DNA and protein sequences has been of interest to evolutionary molecular biologists for decades, but the tools and data available did not allow the fine-grained and nearly complete analyses now possible. For example, one can align sequences for specific genes or regions of interest and catalog all the differences between human and chimpanzee. This in turn allows hypothesis generation regarding the effects of the nucleotide changes on the expression and function of genes and proteins. Traditionally, genomic differences were expected to result in one or more of three general types of human-specific events: changes in the expression patterns or levels of the messenger RNA products of genes (King & Wilson 1975), functionally significant changes in sequences of the protein or noncoding RNAs (Li & Saunders 2005), or major changes resulting in loss-of-function of genes (Olson & Varki 2003, Wang et al. 2006). Multiple examples for each type of event have now been found, making moot the traditional arguments about which is the most "important" mechanism. Examples such as changes in brain gene expression, accelerated changes in gene sequences that may contribute to brain development, and in loss of gene function in protein glycosylation are each detailed below. The studies that led to these findings were each greatly aided by the human and chimpanzee sequencing efforts.

Larger-scale comparisons also allow recognition of genome segments that were lost or duplicated in one species compared with the others. These may not always be recognized in the current state of the chimpanzee draft, because it partly utilized the human sequence to direct its assembly. Regardless, abundant evidence demonstrates significant differences in this class of variation, with many known human duplications not found in the chimpanzee (Fortna et al. 2004, Cheng et al. 2005). Additional work demonstrated more than 200 genes in regions showing copy-number variation or inversion between human and chimpanzee. These large-scale changes could contribute to species differences through loss or gain of gene copies. Analysis of changes in the number and variety of genes in gene families is of particular interest because it offers evolution a relatively rapid avenue to adding and/or changing gene function without altering existing genes. Investigators have proposed that this mechanism accounts for a human-specific expansion of an interesting brain-expressed gene family found on chromosome 16, among others (Johnson et al. 2001). Population-level data can also reveal signatures of recent selection by comparing genetic loci suspected of influencing the current human phenotype. Demonstration that a novel variant in the human has been recently fixed in most or all sectors of the human population is an indication of a strong selective advantage. One example of this is the novel *FOXP2* gene variant found in humans and absent from other primates (see below).

WHAT WILL GENOMIC COMPARISONS NOT TELL US?

As emphasized above, genomic comparisons cannot define functional consequences directly. As an example, the presence of nonsense (stop) or missense (amino acid changing) codons in several genes in the human genome has been detected by comparison with the draft chimpanzee genome. In each instance, the function may have been eliminated or modified in the human and retained in the chimpanzee. An excellent example is the gene encoding Caspase-12, an enzyme normally involved in programmed cell death. It has become inactive in the human (Wang et al. 2006). However, the evolutionary significance of this finding remains speculative,

and the effect of such a modified gene product must be directly tested before any conclusion can be made. Likewise, nucleotide alterations predicted to alter the expression of a gene require validation by direct analysis of expression levels in intact organisms. With nearly 35 million nucleotide differences between the species (not including the tens of millions of nucleotides that are also added or deleted), the functional consequences of each change are not likely to be studied. Instead, candidate genes for likely involvement in human-specific differences must be chosen using a variety of criteria (see below for examples).

One approach to candidate gene prioritization is based on differences in tissue expression patterns. Efforts to compare levels of mRNA expression between human and chimpanzee tissue samples have pointed to such candidate genes (Enard et al. 2002a, Caceres et al. 2003, Gu & Gu 2003, Hsieh et al. 2003, Enard et al. 2004, Khaitovich et al. 2004, Preuss et al. 2004, Uddin et al. 2004). Combining these data with comparative sequence information can suggest DNA alterations underlying expression changes, although this suggestion must be tempered by the potential for expression changes caused by differences in additional, regulatory gene products. Furthermore, the currently available genomic sequences are based on a small number of individuals (or in the case of the chimpanzee, mainly from a single individual at moderate resolution, and a few additional individuals at low resolution). Thus, apparent human-chimpanzee differences may actually represent population variation within a species. This problem has become less acute as additional data have been collected in many human genomes, but the potential remains for polymorphic variation in the chimpanzee population. In this regard, surveying the genomes of a minimum of 10 globally distributed humans can ensure a high probability that a given sequence is fixed (Enard & Paabo 2004). This minimum is likely to be higher for chimpanzees because of their greater level of intraspecific diversity (Gagneux et al. 1999, Kaessmann et al. 1999, Stone et al. 2002, Yu et al. 2003).

POTENTIAL PROBLEMS IN COMPARING TWO CLOSELY RELATED GENOMES

Comparing closely related genomes offers opportunities and also some challenges. One example can be found in the attempt to define "important" or functionally significant sequences. Evolutionary conservation of genomic sequences is usually interpreted as a signature of functional significance owing to selection against loss or change of the sequences. When genomic sequences are distantly related, the sequences that have remained closely related (conserved) are easily identified. Selection against loss or alteration of these sequences is the presumptive mechanism for their conservation. It is common to find that sequences that encode proteins (exons) are more conserved than are intervening sequences within genes (introns) or other nongenic sequences. However, sequences within introns or near conserved genes are often found to retain regulatory or other functions under selection. With the high level of identity found between human and chimpanzee, there is sometimes too much conservation for such functional differences to be clearly defined. This situation requires comparison with additional genomes (see below).

Another technical issue is the use of the reference human sequence to assemble the chimpanzee genome and the draft state of the latter sequence. Sequencing genomes to completion (high resolution definition of all nucleotides at >tenfold redundancy) is currently prohibitively expensive, but a significant fraction of the useful information regarding a genome can be determined by sequencing to draft levels, typically sixfold redundancy. This approach still leaves considerable gaps and regions of uncertain sequence. The chimpanzee genome has been determined at this level, but assemblies have been guided by the human

sequence. This practice can result in errors arising from the assumption that the human and chimpanzee genomes are sufficiently similar as to lack major rearrangements. Finished sequencing is underway for two chimpanzee chromosomes (7 and Y) at the Washington University Genome Sequencing Center, but completion of the chimpanzee sequence to the standards of the current human reference sequence will likely not occur until sequencing costs are trimmed significantly.

IMPORTANCE OF OUT-GROUP GENOMES

Additional genome sequences are important for understanding whether the change occurred on the chimpanzee or on the human lineage because the state common to a more distant species is likely ancestral. In these instances, a not-too-distantly related common ancestor (e.g., orangutan) is best. One should also compare such "out-group" genome sequences to determine the likelihood that a change is functionally important. Here, a somewhat more distantly related out-group species (e.g., rhesus macaque) can be helpful. Such comparisons of gene and protein sequences within primate phylogeny have been done for several decades (Doolittle et al. 1971), and the principles involved have not changed markedly with the introduction of whole genome analyses, with the exception that genome-scale data are now becoming available for many primate species. A draft sequence of the rhesus macaque Old World monkey (*Macaca mulatta*) has just been published (Rhesus Macaque Genome Seq. Anal. Consort. 2007), and orangutan, gorilla, and marmoset New World monkey (*Callithrix jacchus*) genomes are in early stages. The recent announcement of plans to sequence the Northern white-cheeked gibbon (*Nomascus leucogenys*) completes representation of at least one nonhuman primate genome from each of the major lineages along the primate evolutionary tree. With the ability to compare genomes of such species, the interpretation of differences between human and chimpanzee will greatly improve.

GENOME-WIDE VERSUS CANDIDATE GENE APPROACHES

Investigators use two general approaches to utilizing the chimpanzee genome sequence. In the first, one surveys the entire chimpanzee genome and compares it with that of humans, looking for unique features and differences relevant to understanding human evolution. As discussed above, this genome-wide approach can be powerful, and useful generalizations can emerge, especially when combined with analyses of other out-group genomes. However, the large amount of information can also make it difficult to decide where to focus further efforts, or even result in "missing the trees for the forest." At the other extreme, the candidate gene approach involves selecting and studying one or more genes based on known or potential biological differences between humans and chimpanzees. The latter approach can be powerful if gene selection is based on some reasonably direct evidence, but it can also be misleading if based purely on a hunch. An intermediate approach is to explore an entire class of genes predicted to demonstrate significant differences between humans and chimpanzees. Some examples of each of these approaches are presented below.

EXAMPLES OF PROMISING RESULTS FROM GENOME-WIDE APPROACHES

Examination of genomic differences between humans and chimpanzees began when Yunis and coworkers explored differences in chromosome banding patterns between humans and great apes (Yunis & Prakash 1982), defining large lineage-specific chromosome rearrangements. The most obvious rearrangement was the fusion resulting in human chromosome 2, reducing the human chromosome number relative to the other hominid

species (Fan et al. 2002). Inversions of large segments within chromosomes were the other major events, and most likely occurred on the chimpanzee lineage (Eichler et al. 1996, Nickerson & Nelson 1998). All these "pericentric" inversions (large chromosomal rearrangements in which the center portion of the chromosome containing the centromere inverts between long and short chromosome arms) are also found in the bonobo (Kehrer-Sawatzki & Cooper 2006), which shared a common ancestor with a chimpanzee ~2 Mya. Thus the possibility arises that these events were related to the original separation of the human and chimpanzee/bonobo lineages from a common ancestor and/or to the emergence of uniquely human features.

Several groups pursued the regions involved in the breakage and rejoining events that flank these large-scale alterations. The underlying hypothesis was that rearrangements may have disrupted and/or altered the patterns of expression of genes at or near the breakpoint regions. However, each of the regions has now been investigated, and no evidence for gene disruption or major alterations in gene expression has been found so far. Of interest has been the identification, in these same regions, of "low copy repeats" (LCRs), which are duplicated sequences specific to the species where the rearrangement occurred. These could potentially contribute to gene regulation or copy number changes. The LCRs associated with gross chromosome scale rearrangements turn out to represent a very small fraction of all LCRs, which have been recently recognized as a significant source of variation within and between species because they can show significant differences in copy number (Redon et al. 2006). This level of variation is likely to support major alterations in gene copy number and therefore expression, leading in principle to species-specific differences. Cataloging this variation in human genome polymorphism and disease, as well as in comparative genomic studies, is currently a very active area of investigation (Khaja et al. 2006, Locke et al. 2006, Redon et al. 2006) and has led to such interesting discoveries as polymorphic variation in susceptibility to infection by HIV (Gonzalez et al. 2005).

Similar comparative studies of the chimpanzee genome are somewhat hampered by the current state of the sequence data, and current estimates of chimpanzee genome LCR content must be interpreted with caution owing to the draft nature of the sequence and to the lack of population data (Perry et al. 2006). Nonetheless, confirmed data do support differences in sequence composition between humans and chimpanzees of at least 24 million nucleotides when relatively large segments (more than 15,000 base pairs) are studied, including significant losses or gains in sequence (Newman et al. 2005). Smaller regions with changes in sequence content are also present but are more difficult to characterize with the draft chimpanzee sequence. Some 840,000 small (fewer than 100 base pairs) insertions and deletions in the human were detected after comparison with multiple species including the chimpanzee (Chen et al. 2006). Many of these were found to affect known protein-coding genes in the human.

Sequence alterations are often caused by mobile elements (so-called jumping genes), which comprise almost half of the human genome. Many classes of these "transposable" elements are found in primate genomes, although the vast majority of these are no longer capable of mobilization (i.e., jumping into new sites in the genome). Although ongoing transposition is supported by a small number (tens to hundreds, depending on the type of transposon) of active copies, most transposed elements are relics of earlier events. Mills and coworkers (2006) found nearly 11,000 transposable element insertions that occurred independently in the human and chimpanzee lineages. Roughly one-third of the insertions were found within known gene regions, with the potential to alter gene expression or function. In one class of transposon (Alu sequences), the number of mobilized elements has been accelerated some threefold

in the human compared with the chimpanzee (Hughes et al. 2005). Within humans, one can find well-characterized examples of transposon insertions that functionally alter the genome. Examples include the inactivation of the *CMP-N-acetylneuraminic acid hydroxylase* (*CMAH*) (Hayakawa et al. 2001) and *SIGLEC13* genes (Angata et al. 2004) in the human lineage (see below). Other transposition events are polymorphic within the human species, indicating a recent origin (Cordaux et al. 2006). Although events that are human- or chimpanzee-specific have the potential to contribute to species differences, the large numbers will require considerable effort to determine effects on local gene function(s).

GENE FAMILIES AND SELECTIVE PRESSURES

Although gene families (groups of genes derived from ancient or recent duplication events) are typically found in clusters, members can be distributed to other chromosomal locations. Expansion and contraction of gene numbers in such clusters have been postulated as drivers of evolutionary change (Johnson et al. 2001, Angata et al. 2004, Birtle et al. 2005) because new gene copies are presumably free to acquire new functions, whereas the original copy carries on the primary function that was under selective pressure. An excellent human-specific example is a gene family (MGC8902) amplified in the human lineage (Popesco et al. 2006). This candidate was identified, using a microarray approach, after a search for sequences increased in humans relative to chimpanzees (Pollack et al. 1999) and was found to be among the most dramatically increased in copy number in humans. The family is also polymorphic in copy number among humans, ranging between 40 and 65 copies, compared with fewer than 10 in chimpanzee and bonobo and between 10 and 15 in gorilla (Popesco et al. 2006). Copy number in Old World monkeys is much lower, and other mammals appear to have a single copy. Within copies of MGC8902 are multiple occurrences of a two-exon segment encoding a protein domain called "DUF1220," a 65-amino-acid sequence of unknown function. Genes containing coding sequences for DUF1220 domains are highly expressed in neurons, both in cell bodies and dendrites. Of further interest is the finding that such genes also exhibit strong signatures for positive selection (Popesco et al. 2006). These signatures are detected by comparing the ratio of nonsynonymous amino acid changing nucleotide sequences to synonymous (nonamino acid changing) ones (the so-called Ka/Ks ratio). A ratio above 1.0 is traditionally considered a signature of positive selection,[1] and some one-third of human DUF1220 domain sequences have ratios above 1.0. The data suggest expansion and selection of this gene family in primates that were greatly accelerated in the human lineage. Although the function(s) of these genes remains to be determined, these genes appear to be contenders for conferring changes on neuronal function.

Another example is the olfactory receptor (OR) gene family, comprising ~1000–1400 members in vertebrate genomes, which is responsible for the complexities of the sense of smell. A significant fraction can be defective in a given species (e.g., ~20% in the mouse). One striking feature of the human genome is that ~60% of the >1000 OR genes are nonfunctional. A significantly higher percentage of chimpanzee OR genes has been maintained (Gilad et al. 2005), suggesting that humans came to rely less on the sense of smell, allowing mutations to accumulate under relaxed selective constraints. In contrast, smaller

[1]A Ka/Ks ratio above 1.0 is traditionally considered a signature of positive selection. However, lower numbers can be meaningful. For example, whereas some amino acids of a protein are under positive selection, others remain under purifying selection, which constrains changes. Thus, the average Ka/Ks ratio of the entire protein can be misleadingly low. In some instances where proteins have distinct domains, one can define positive selection in one domain, in comparison to adjacent ones (Angata et al. 2004). Some sophisticated methods can even be used to define selection rates at individual amino acid positions (Yang et al. 2000).

subsets of OR genes have expanded in each species and appear to be under positive selection, with some being polymorphic (Menashe et al. 2003).

We can now compare sequences across the whole human and chimpanzee genomes for indications of selection. Such a scan recently defined 49 regions where the sequences are conserved among mammals but are rapidly diverging in humans compared with chimpanzees (Pollard et al. 2006). These were defined as human accelerated regions (HARs) and were enumerated from 1 to 49 by their order of significance, as measured by the extent of the change in the human away from the conserved consensus mammalian sequence. All but one was not a protein coding sequence. The top candidate showing the most human-specific changes (HAR1 on chromosome 20) underwent accelerated changes ~20-fold above expectations for random events. The HAR1 locus does not contain protein-coding sequences but does, however, make RNA transcripts. One of these (HAR1F) is expressed in neurons with a role in early development of the neocortex, a brain region associated with higher cognitive functions. This suggests a role for the HAR1 locus in brain development, and possibly in differences unique to human brain development. With the growing appreciation that RNAs can function in many regulatory roles in addition to their previously known functions (Kloosterman & Plasterk 2006), species-specific changes may be more easily conferred through changes of regulatory RNAs, which could affect many targets. This idea is a variation of the notion that regulatory changes might supercede alterations of individual proteins in generating species differences (King & Wilson 1975).

EXAMPLES OF PROMISING RESULTS FROM CANDIDATE GENE APPROACHES

Just over a decade ago there were only a handful of genes with known differences between humans and chimpanzees, and it was unclear whether some of these were actually human-universal and human-specific. The situation has changed dramatically, and we now face an "embarrassment of riches," with many candidate genes vying for attention. Space does not allow a discussion of all of them. A recent listing can be found in Varki & Altheide (2005), and a few examples are discussed below.

CMAH

The first human-specific genetic difference with a clear-cut biochemical outcome was an inactivating defect in the *CMAH* gene (Chou et al. 1998, Irie et al. 1998, Chou et al. 2002). This gene encodes an enzyme (cytidine monophosphate N-acetylneuraminic acid hydroxylase) that converts one form of sialic acid, N-acetylneuraminic acid (Neu5Ac), to another, called N-glycolylneuraminic acid (Neu5Gc). Because sialic acids are attached to the surfaces of every cell in the body, this human-specific defect generates a marked difference in the human "sialome" (the sum total of all sialic acids and their presentation on cell surfaces and secreted molecules) from that of other hominids—a loss of Neu5Gc and an excess of the precursor Neu5Ac (Muchmore et al. 1998). This has multiple implications for the human phenotype, ranging from resistance and susceptibility to infections, to antigenic responses to animal products, to secondary effects on the immune response, and to potential effects on brain development (Varki 2001, 2007). The human *CMAH* mutation apparently occurred ~2–3 Mya, according to analysis of sialic acids in fossils and molecular analysis of the genetic locus (Chou et al. 2002). This occurred just prior to the emergence of the genus *Homo* (Wood & Collard 1999). However, although Neanderthals (with whom humans shared a common ancestor about half a million years ago) were also lacking Neu5Gc (Chou et al. 2002), there is no direct evidence that this was also the genotype of the ancestral *Homo ergaster/erectus*.

FOXP2

Studies of a family with an inherited speech production disorder resulted in isolation of a gene defective in all affected family members (Lai et al. 2001, Vargha-Khadem et al. 2005). This gene, called *FOXP2*, encodes a known transcription factor (a protein that binds to DNA and changes gene expression). Although the normal gene was practically identical between the chimpanzee and the mouse, humans have two amino acid changes in a functional domain of the molecule (Enard et al. 2002b, Zhang et al. 2002). Furthermore, analysis of the human gene suggested that these changes occurred within the past 200,000 years (and most likely more recently, in the time immediately preceding the large growth of human populations). Although this is not a "language gene" (as proclaimed by the media) it does seem to be involved in proper articulation of human speech and possibly in some aspects of grammar generation (Vargha-Khadem et al. 2005). This may be one of the genes in which specific changes occurring in the human lineage contributed to the human phenotype. Indeed, mice with a disruption in a single copy of this gene showed a modest developmental delay and a significant alteration in ultrasonic vocalizations normally elicited when pups are removed from their mothers (Shu et al. 2005). The next step is to replace the mouse gene with the human version and evaluate what changes occur.

ASPM and *MCPH1*

ASPM and *MCPH1* are genes that when defective result in the human syndrome of microcephaly (i.e., a small head and brain). Given the unusual expansion of the brain during human evolution (Wood & Collard 1999), it was reasonable to suggest that these genes contributed to the human phenotype. Indeed, examination of both genes shows evidence of recent accelerated evolution and positive selection (Zhang 2003; Evans et al. 2004a,b; Kouprina et al. 2004). Of course it is difficult to be sure that changes in these genes are responsible for the human phenotype. In this regard, an intriguing finding is evidence for single alleles that have recently swept through the human population (Evans et al. 2005, Mekel-Bobrov et al. 2005), one possibly originating even from interbreeding with archaic *Homo sapiens* (Evans et al. 2006). Again, modeling the human genotype in a mouse could help address the issue.

Siglecs

This family of cell surface molecules recognizes different types and attachments of sialic acids, with expression prominent in the immune system. The ancestral functional state of hominid Siglecs appears to have been to recognize Neu5Gc (Sonnenburg et al. 2004). Thus, loss of Neu5Gc expression during human evolution (due to the *CMAH* mutation, see above) likely resulted in a sudden loss of binding sites for several Siglecs (Varki 2007). This might have resulted in a "hyperimmune" state, which could have been advantageous in the short run but may have eventually precipitated substantial changes in the patterns and expression of Siglecs. For example, whereas all other hominids express multiple Siglecs on their T-lymphocytes, humans suppressed this expression. This loss of negative signaling may account for the hyperreactivity of human T cells (Nguyen et al. 2006). Also, a human-specific "gene conversion" of *SIGLEC11* occurred, in which the front portion of the gene was replaced by a sequence from a nearby pseudogene (Hayakawa et al. 2005). Normally such events result in inactivation of the "converted" gene as well. Instead, this event maintained a functional gene that encodes a truly human-specific protein. It also changed the binding specificity of the encoded Siglec-11 protein and resulted in new expression in brain microglia. The significance of this human-specific event is unknown, but it could reflect susceptibility or resistance to sialylated pathogens that enter the brain and/or effects on the roles of microglia in regulating aspects of neural function.

Additional human-specific changes have been found in the Siglecs (Brinkman-Van der Linden et al. 2000, Angata et al. 2001, Varki 2007). Interpreting these changes must be done cautiously because some Siglecs appear to be rapidly evolving in all taxa where they have been studied. However, in every instance studied, comparison with other "great ape" out-groups indicates that the human-chimpanzee difference is specific to humans (Varki 2007).

EXAMPLES OF PROMISING RESULTS FROM COMBINED APPROACHES

In contrast to broad genome-wide surveys and candidate gene analyses, a combined approach considers general clues about unique aspects of the human phenotype and surveys all genes potentially related to the biology of the involved system.

Skin and Appendages

Although the brain is assumed to be the organ showing the most significant changes between humans and great apes, this may not be the case. The biology of the skin and its appendages appears to have undergone more major changes (Montagna & Yun 1963, Whitford 1976), such as loss of hair, gain of subcutaneous fat, changes in sweat glands, and the development of fully formed female breasts (which are a specialized type of skin gland) (Oftedal 2002), without the usual mammalian stimulus of pregnancy and lactation. Humans have lost one of the keratin genes involved in the development of hair, and the protein encoded by this gene is expressed in an unusual pattern in chimpanzee hair (Winter et al. 2001). However, no direct connection has been made to the obvious differences in the biology and display of hair between humans and great apes. Additionally, changes have been noted in clusters of genes involved in skin development (Chimpanzee Seq. Anal. Consort. 2005). The field needs a systematic analysis of all genes involved in the development and maintenance of skin.

Brain

Researchers took two types of approaches with regard to the brain. The first looked for unique patterns of gene expression in the brain, by comparative studies of messenger RNA expression (Enard et al. 2002a, Caceres et al. 2003, Gu & Gu 2003, Enard et al. 2004, Khaitovich et al. 2004, Preuss et al. 2004, Uddin et al. 2004). Although these studies have yielded many tantalizing clues, no single candidate gene emerged, and some controversy about interpretations has developed. A parallel effort focused on a collection of several hundred genes known to be involved in brain development, making comparisons between humans, monkeys, mice, and rats (Dorus et al. 2004). In comparison with other tissues, the genes involved in brain development did indeed showed more evidence for accelerated changes and evidence of positive selection in the human lineage.

Muscles

Another unexplained difference between humans and great apes is the apparent loss of muscle strength in humans. A myosin gene prominently expressed in jaw muscles is inactivated in humans, apparently ~2 Mya (Stedman et al. 2004) (although not all investigators agree with this timing) (Perry et al. 2005). The original authors speculated that loss of this gene could explain the marked reduction in the size and area of attachment of the jaw muscles in humans and that this reduction of muscles permitted evolution of a large brain, by relaxing physical constraints on cranial expansion. Although the first idea remains plausible, the second appears a bit far-fetched, because so many other factors determine the development of a large brain, including the expression of genes mentioned above and the delay in closure of the cranial sutures that bind together the bones of the human skull

(Cohen 1991). As human ancestors switched to a less herbivorous diet, the loss of *MHY16* and jaw muscle strength might have been inconsequential and simply allowed the mutation to drift to fixation (Currie 2004).

Reproductive Biology

Another system showing striking differences between humans and apes is the reproductive tract. Apart from the marked differences in breast development mentioned above, other differences can be found, such as the massive skin swellings of fertile female chimps, the different structure and size of the penis, and the relative size of the testes (Harcourt et al. 1981). Human female reproduction is also characterized by loss of obvious signals regarding the timing of ovulation (Nunn 1999) and an apparent increase in menstrual blood loss, which may explain the high frequency of iron deficiency anemia in the human female (Newman 2006). These and other changes suggest that studies of reproductive biology genes should be fruitful; however, to date these studies have been limited.

Immunity

Humans and great apes display multiple immunological differences, such as the Siglec changes and hyperreactivity of T lymphocytes mentioned above, the apparently high frequency of autoimmune disease (Varki 2000), and the loss of the *CASPASE12* gene (Wang et al. 2006). Although a systematic survey is worthwhile, one must remember that the immune system is rapidly evolving in all taxa and that some differences are to be expected.

RELATING GENOMIC TO PHENOMIC DIFFERENCES IN THE CONTEXT OF ENVIRONMENTAL INFLUENCES

Even with complete high-resolution genomes of humans, chimpanzees, other hominids, and other primate species, it will be difficult to derive meaningful conclusions about human evolution without considering the relevant phenotypic differences and the impact of the environment on generating these differences. As discussed above, paucity of information regarding the "phenome" of the other living hominids is remarkable. [The term phenome has been used in many publications—e.g., Mahner & Kary (1997), Varki et al. (1998), Paigen & Eppig (2000), Nevo (2001), Walhout et al. (2002), Freimer & Sabatti (2003)—but still lacks a universally accepted definition. Discussions with others who have used the term suggest the following definition: The body of information describing an organism's phenotypes, under the influences of genetic and environmental factors.]

Thus, regardless of how much information we have about the human phenome (Freimer & Sabatti 2003), we will be unable to relate the genomic differences to features of the human phenotype without equivalent information on the phenome of the great apes. We and others have therefore suggested a great ape phenome project, in which the phenotypic features of the great apes would be cataloged and compared with those of humans, accounting for effects of other factors, such as the environment (Varki et al. 1998, McConkey & Varki 2005). From the point of view of understanding human origins, it makes sense to focus first on systems wherein there appear to be major differences between humans and other hominids, such as in the nervous, immune, and reproductive systems, in the skin, etc. However, such a focused approach might miss other differences. We need a catalog of all differences not only in normal physiology and homeostasis, but also in diseases, where there appear to be some marked disparities (Varki 2000, Olson & Varki 2003, Varki & Altheide 2005). Studies of these differences could help us understand unique features of the human phenotype.

Such phenotypic data rely on studies of great apes both in captivity and in the wild. Each setting has its advantages and

disadvantages. Although some feel that studies done outside the "natural condition" are irrelevant, this is not necessarily the case. After all, we humans are no longer living in our "natural" environment, and comparisons with great apes living in enclosed spaces may actually be relevant because we have "self-domesticated" and also live in enclosed spaces. Studying captive great apes also offers better access to physiological, biochemical, cellular, and molecular information that emerges from their medical care. Unfortunately, most great apes other than chimpanzees have been moved to small groups in zoos, where such information is collected to a limited extent and/or is not easily accessible. Meanwhile, populations of captive chimpanzees in the United States are rapidly dwindling because of a government-enforced moratorium on breeding (VandeBerg & Zola 2005, Cohen 2007). This moratorium originated not only in fiscal issues, but also in the realization that chimpanzees are no longer considered by many to be ethically suitable subjects for invasive experimentation (Gagneux et al. 2005, Cohen 2007). It has nevertheless served to stifle all forms of research aimed at understanding chimpanzees, even via approaches that most investigators would consider ethical. Finally, there is limited use in having the genomic DNA sequence of a chimpanzee if one does not also know which RNAs, proteins, lipids, and glycans are being expressed in which tissues, and at what stage of development. This research could be approached with noninvasive studies using easily accessible materials such as the blood, and also via autopsy, when chimpanzees die of natural causes or are euthanized because of terminal suffering.

ETHICAL, LEGAL, AND SOCIAL ISSUES

The Human Genome Project was accompanied by funded programs in ethical, legal, and social issues. This was not done, however, for the Chimpanzee Genome Project. Although the issues are less complex, concerns have arisen (Gagneux et al. 2005). For example, should this information be used to design genetic experiments in apes, such as selective breeding, transgenic modification, or gene knockouts? The answer should surely be no. Second, should this information provide an impetus for markedly increased research on chimpanzees and other great apes? The answer here is a guarded yes, with a proviso that the types of experiments done should probably follow ethical principles generally similar to those used in the study of humans. After all, there is limited value in doing a study on a chimpanzee if the results cannot be compared with a similar study in humans. In reality, there is a wide spectrum of opinions about the ethical issues related to the management and utility of chimpanzees in captivity (Brent 2004, Gagneux et al. 2005, VandeBerg & Zola 2005, Cohen 2007). While this debate is ongoing, major funding agencies have unfortunately either banned all research on captive chimpanzees (as in many European countries) or halted all breeding (as in the United States). It is fitting to ask what would be the reaction to a complete ban on all research on human subjects. Worse still, the moratorium will eventually result in decimating the population of captive chimpanzees in the United States (Cohen 2007). This would be a sad loss, especially if there is a parallel loss of the chimpanzee population in the wild because of other unrelated human activities, such a deforestation and bush-meat trading.

FUTURE PROSPECTS

Sequencing of the chimpanzee genome raised a great deal of expectations. In fact, although it is a powerful resource and an extremely useful database, it is also nothing more than a "parts list." Furthermore, the number of genomic differences between the human and chimpanzee genome is quite substantial, and much work needs to be done before we even begin to approach the original question of anthropogeny, i.e., explaining the human phenotype and the origin of humans. Regardless, with the

ongoing sequencing of several other primate genomes one can anticipate rapid progress.

Even if a single gene or set of genes is clearly different between humans and great apes and the data are completely consistent with some aspects of a biological difference, it remains a challenge to prove the case conclusively. One approach is to discover a human mutant in the relevant gene. However, it is not obvious how to search for such human mutants because the effects of gene dysfunction or loss might be much more complex than predicted by the hypothesis arising from the study of a gene. A second approach is to deliberately model a specific human genotype in animals. To do this in any great ape would be not only impractical and enormously expensive, but unethical in the minds of most scientists. Great challenges exist in carrying out transgenic studies in other primates because this technology is very poorly developed, and the generation time of most primates is too long to complete such studies within the scientific lifetime of an investigator. The mouse then is the currently optimal system in which to carry out genetic manipulations because the technology is well developed and continues to be feasible. Of course, a mouse is not a primate, and mimicking the human genotype in the mouse does not recapitulate what would have happened if the same genetic event had occurred in the hominid lineage. Regardless, this seems at the moment to be the most practical and rational way to proceed.

Finally, even if we had an extensive catalog of specific genetic changes that occurred in the human lineage, and could find reasonable evidence from human mutants or induced mouse mutants to support resulting hypotheses, the story will remain incomplete until we know the timing of occurrence and fixation of each genetic change during human evolution. This has been achieved only in a few instances, taking advantage of studies of fossils or molecular fossils, i.e., the inactivated genetic locus. Examples include the timing of inactivation of the *CMAH* gene (done by studying sialic acids in fossils and the sequences of the inactivation region) (Chou et al. 2002), and the *FOXP2* gene (done by comparing genomes of human populations) (Enard et al. 2002b, Zhang et al. 2002). Indeed, if all genetic differences between humans and chimpanzees specific to the human lineage were defined and the timing of each of these genetic events could be established within the past six to seven million years, the story would almost tell itself. As such a database is developed it will also be important to address currently popular "umbrella" theories that claim to explain many aspects of human uniqueness with a single cause or biological process (Langdon 1997). Any such analysis must account for these abovementioned issues of timing.

The evolutionary origin of humans is not likely to be fully explained within our lifetimes. Nevertheless, the prospect of being able to approach this grand challenge using genomic information is truly exciting, and investigators will probably find many interesting surprises ahead. To use a clichéd phrase, this is "not the beginning of the end, but the end of the beginning" in this great adventure.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Angata T, Margulies EH, Green ED, Varki A. 2004. Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. *Proc. Natl. Acad. Sci. USA* 101:13251–56

Angata T, Varki NM, Varki A. 2001. A second uniquely human mutation affecting sialic acid biology. *J. Biol. Chem.* 276:40282–87

Birtle Z, Goodstadt L, Ponting CP. 2005. Duplication and positive selection among hominin-specific PRAME genes. *BMC Genomics* 6:120

Brent L. 2004. Solutions for research chimpanzees. *Lab. Anim. (NY)* 33:37–43

Brinkman-Van der Linden ECM, Sjoberg ER, Juneja LR, Crocker PR, Varki N, Varki A. 2000. Loss of N-glycolylneuraminic acid in human evolution—implications for sialic acid recognition by siglecs. *J. Biol. Chem.* 275:8633–40

Britten RJ. 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl. Acad. Sci. USA* 99:13633–35

Brunet M, Guy F, Pilbeam D, Mackaye HT, Likius A, et al. 2002. A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* 418:145–51

Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, et al. 2003. Elevated gene expression levels distinguish human from nonhuman primate brains. *Proc. Natl. Acad. Sci. USA* 100:13030–35

Carroll SB. 2003. Genetics and the making of *Homo sapiens*. *Nature* 422:849–57

Chen FC, Chen CJ, Li WH, Chuang TJ. 2007. Human-specific insertions and deletions inferred from mammalian genome sequences. *Genome Res.* 17:16–22

Chen FC, Li WH. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* 68:444–56

Cheng Z, Ventura M, She X, Khaitovich P, Graves T, et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* 437:88–93

Chimpanzee Seq. Anal. Consort. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87

Chou HH, Hayakawa T, Diaz S, Krings M, Indriati E, et al. 2002. Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl. Acad. Sci. USA* 99:11736–41

Chou HH, Takematsu H, Diaz S, Iber J, Nickerson E, et al. 1998. A mutation in human CMP-sialic acid hydroxylase occurred after the *Homo-Pan* divergence. *Proc. Natl. Acad. Sci. USA* 95:11751–56

Cohen J. 2007. Biomedical research. The endangered lab chimp. *Science* 315:450–52

Cohen MMJ. 1991. Etiopathogenesis of craniosynostosis. *Neurosurg. Clin. N. Am.* 2:507–13

Consort. IHM. 2005. A haplotype map of the human genome. *Nature* 437:1299–320

Cordaux R, Hedges DJ, Herke SW, Batzer MA. 2006. Estimating the retrotransposition rate of human Alu elements. *Gene* 373:134–37

Currie P. 2004. Human genetics: muscling in on hominid evolution. *Nature* 428:373–74

Doolittle RF, Wooding GL, Lin Y, Riley M. 1971. Hominoid evolution as judged by fibrinopeptide structures. *J. Mol. Evol.* 1:74–83

Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, et al. 2004. Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* 119:1027–40

Eichler EE, Lu F, Shen Y, Antonacci R, Jurecic V, et al. 1996. Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum. Mol. Genet.* 5:899–912

Enard W, Fassbender A, Model F, Adorjan P, Paabo S, Olek A. 2004. Differences in DNA methylation patterns between humans and chimpanzees. *Curr. Biol.* 14(4):R148–49

Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, et al. 2002a. Intra- and interspecific variation in primate gene expression patterns. *Science* 296:340–43

Enard W, Paabo S. 2004. Comparative primate genomics. *Annu. Rev. Genomics Hum. Genet.* 5:351–78

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, et al. 2002b. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418:869–72

Evans PD, Anderson JR, Vallender EJ, Choi SS, Lahn BT. 2004a. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. *Hum. Mol. Genet.* 13:1139–45

Evans PD, Anderson JR, Vallender EJ, Gilbert SL, Malcom CM, et al. 2004b. Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. *Hum. Mol. Genet.* 13:489–94

Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, et al. 2005. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* 309:1717–20

Evans PD, Mekel-Bobrov N, Vallender EJ, Hudson RR, Lahn BT. 2006. Evidence that the adaptive allele of the brain size gene microcephalin introgressed into *Homo sapiens* from an archaic *Homo* lineage. *Proc. Natl. Acad. Sci. USA* 103:18178–83

Fan Y, Linardopoulou E, Friedman C, Williams E, Trask BJ. 2002. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13–2q14.1 and paralogous regions on other human chromosomes. *Genome Res.* 12:1651–62

Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, et al. 2004. Lineage-specific gene duplication and loss in human and great ape evolution. *PLOS Biol.* 2:E207

Freimer N, Sabatti C. 2003. The human phenome project. *Nat. Genet.* 34:15–21

Gagneux P. 2004. A *Pan*-Oramic view: insights into hominid evolution through the chimpanzee genome. *Trends Ecol. Evol.* 19:571–76

Gagneux P, Moore JJ, Varki A. 2005. The ethics of research on great apes. *Nature* 437:27–29

Gagneux P, Varki A. 2001. Genetic differences between humans and great apes. *Mol. Phylogenet. Evol.* 18:2–13

Gagneux P, Wills C, Gerloff U, Tautz D, Morin PA, et al. 1999. Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc. Natl. Acad. Sci. USA* 96:5077–82

Gilad Y, Man O, Glusman G. 2005. A comparison of the human and chimpanzee olfactory receptor gene repertoires. *Genome Res.* 15:224–30

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, et al. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307:1434–40

Goodman M. 1999. The genomic record of Humankind's evolutionary roots. *Am. J. Hum. Genet.* 64:31–39

Goodman M, Grossman LI, Wildman DE. 2005. Moving primate genomics beyond the chimpanzee genome. *Trends Genet.* 9:511–17

Gu J, Gu X. 2003. Induced gene expression in human brain after the split from chimpanzee. *Trends Genet.* 19:63–65

Harcourt AH, Harvey PH, Larson SG, Short RV. 1981. Testis weight, body weight and breeding system in primates. *Nature* 293:55–57

Hayakawa T, Angata T, Lewis AL, Mikkelsen TS, Varki NM, Varki A. 2005. A human-specific gene in microglia. *Science* 309:1693

Hayakawa T, Satta Y, Gagneux P, Varki A, Takahata N. 2001. Alu-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. *Proc. Natl. Acad. Sci. USA* 98:11399–404

Hsieh WP, Chu TM, Wolfinger RD, Gibson G. 2003. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* 165:747–57

Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, et al. 2005. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* 437:100–3

Int. Hum. Genome Seq. Consort. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931–45

Irie A, Koyama S, Kozutsumi Y, Kawasaki T, Suzuki A. 1998. The molecular basis for the absence of N-glycolylneuraminic acid in humans. *J. Biol. Chem.* 273:15866–71

Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, et al. 2001. Positive selection of a gene family during the emergence of humans and African apes. *Nature* 413:514–19

Kaessmann H, Wiebe V, Paabo S. 1999. Extensive nuclear DNA sequence diversity among chimpanzees. *Science* 286:1159–62

Kehrer-Sawatzki H, Cooper DN. 2006. Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. *Hum. Mutat.* 28:99–130

Khaitovich P, Muetzel B, She X, Lachmann M, Hellmann I, et al. 2004. Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* 14:1462–73

Khaja R, Zhang J, Macdonald JR, He Y, Joseph-George AM, et al. 2006. Genome assembly comparison identifies structural variants in the human genome. *Nat. Genet.* 38:1413–18

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–16

Klein J, Takahata N. 2002. Where do we come from? The molecular evidence for human descent. *Hum Genet.* 111:579–81

Kloosterman WP, Plasterk RH. 2006. The diverse functions of microRNAs in animal development and disease. *Dev. Cell* 11:441–50

Kouprina N, Pavlicek A, Mochida GH, Solomon G, Gersch W, et al. 2004. Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. *PLOS Biol.* 2:653–63

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* 413:519–23

Langdon JH. 1997. Umbrella hypotheses and parsimony in human evolution: a critique of the Aquatic Ape Hypothesis. *J. Hum. Evol.* 33:479–94

Li WH, Saunders MA. 2005. News and views: the chimpanzee and us. *Nature* 437:50–51

Locke DP, Sharp AJ, McCarroll SA, McGarth SD, Newman TL, et al. 2006. Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am. J. Hum. Genet.* 79:275–90

Mahner M, Kary M. 1997. What exactly are genomes, genotypes and phenotypes? And what about phenomes? *J. Theor. Biol.* 186:55–63

McBrearty S, Jablonski NG. 2005. First fossil chimpanzee. *Nature* 437:105–8

McConkey EH, Fouts R, Goodman M, Nelson D, Penny D, et al. 2000. Proposal for a human genome evolution project. *Mol. Phylogenet. Evol.* 15:1–4

McConkey EH, Varki A. 2000a. A primate genome project deserves high priority. *Science* 289:1295–96

McConkey EH, Varki A. 2000b. Examining priorities for a primate genome project—response. *Science* 290:1505

McConkey EH, Varki A. 2005. Thoughts on the future of great ape research. *Science* 309:1499–501

Mekel-Bobrov N, Gilbert SL, Evans PD, Vallender EJ, Anderson JR, et al. 2005. Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. *Science* 309:1720–22

Menashe I, Man O, Lancet D, Gilad Y. 2003. Different noses for different people. *Nat. Genet.* 34:143–44

Mills RE, Bennett EA, Iskow RC, Luttig CT, Tsui C, et al. 2006. Recently mobilized transposons in the human and chimpanzee genomes. *Am. J. Hum. Genet.* 78:671–79

Montagna W, Yun JS. 1963. The skin of primates. X. The skin of the chimpanzee (*Pan Satyrus*). *Am. J. Phys. Anthropol.* 21:189–203

Muchmore EA, Diaz S, Varki A. 1998. A structural difference between the cell surfaces of humans and the great apes. *Am. J. Phys. Anthropol.* 107:187–98

Nevo E. 2001. Evolution of genome-phenome diversity under environmental stress. *Proc. Natl. Acad. Sci. USA* 98:6233–40

Newman B. 2006. Iron depletion by whole-blood donation harms menstruating females: the current whole-blood-collection paradigm needs to be changed. *Transfusion* 46:1667–81

Newman TL, Tuzun E, Morrison VA, Hayden KE, Ventura M, et al. 2005. A genome-wide survey of structural variation between human and chimpanzee. *Genome Res.* 15:1344–56

Nguyen DH, Hurtado-Ziola N, Gagneux P, Varki A. 2006. Loss of Siglec expression on T lymphocytes during human evolution. *Proc. Natl. Acad. Sci. USA* 103:7765–70

Nickerson E, Nelson DL. 1998. Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees. *Genomics* 50:368–72

Nunn CL. 1999. The evolution of exaggerated sexual swellings in primates and the graded-signal hypothesis. *Anim. Behav.* 58:229–46

Oftedal OT. 2002. The mammary gland and its origin during synapsid evolution. *J. Mammary Gland Biol. Neoplasia* 7:225–52

Olson MV, Varki A. 2003. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat. Rev. Genet.* 4:20–28

Paigen K, Eppig JT. 2000. A mouse phenome project. *Mamm. Genome* 11:715–17

Perry GH, Tchinda J, McGarth SD, Zhang J, Picker SR, et al. 2006. Hotspots for copy number variation in chimpanzees and humans. *Proc. Natl. Acad. Sci. USA* 103:8006–11

Perry GH, Verrelli BC, Stone AC. 2005. Comparative analyses reveal a complex history of molecular evolution for human MYH16. *Mol. Biol. Evol.* 22(3):379–82

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, et al. 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 23:41–46

Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, et al. 2006. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443:167–72

Popesco MC, Maclaren EJ, Hopkins J, Dumas L, Cox M, et al. 2006. Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* 313:1304–7

Preuss TM, Caceres M, Oldham MC, Geschwind DH. 2004. Human brain evolution: insights from microarrays. *Nat. Rev. Genet.* 5:850–60

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. 2006. Global variation in copy number in the human genome. *Nature* 444:444–54

Rhesus Macaque Genome Seq. Anal. Consort. 2007. Human-specific changes of genome structure detected by genomic triangulation. *Science* 316:235–37

Ruvolo M. 2004. Comparative primate genomics: the year of the chimpanzee. *Curr. Opin. Genet. Dev.* 14:650–56

Shu W, Cho JY, Jiang Y, Zhang M, Weisz D, et al. 2005. Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proc. Natl. Acad. Sci. USA* 102:9643–48

Sonnenburg JL, Altheide TK, Varki A. 2004. A uniquely human consequence of domain-specific functional adaptation in a sialic acid-binding receptor. *Glycobiology* 14:339–46
Stedman HH, Kozyak BW, Nelson A, Thesier DM, Su LT, et al. 2004. Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature* 428:415–18
Stone AC, Griffiths RC, Zegura SL, Hammer MF. 2002. High levels of Y-chromosome nucleotide diversity in the genus *Pan*. *Proc. Natl. Acad. Sci. USA* 99:43–48
Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, et al. 2003. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 424:788–93
Uddin M, Wildman DE, Liu G, Xu W, Johnson RM, et al. 2004. Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc. Natl. Acad. Sci. USA* 101:2957–62
VandeBerg JL, Williams-Blangero S, Dyke B, Rogers J. 2000. Examining priorities for a primate genome project. *Science* 290:1504–5
VandeBerg JL, Zola SM. 2005. A unique biomedical resource at risk. *Nature* 437:30–32
Vargha-Khadem F, Gadian DG, Copp A, Mishkin M. 2005. FOXP2 and the neuroanatomy of speech and language. *Nat. Rev. Neurosci.* 6:131–38
Varki A. 2000. A chimpanzee genome project is a biomedical imperative. *Genome Res.* 10:1065–70
Varki A. 2001. Loss of N-glycolylneuraminic acid in humans: mechanisms, consequences and implications for hominid evolution. *Am. J. Phys. Anthropol.* 33(Suppl.):54–69
Varki A. 2007. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature*. In press
Varki A, Altheide TK. 2005. Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.* 15:1746–58
Varki A, Wills C, Perlmutter D, Woodruff, Gage F, et al. 1998. Great ape phenome project? *Science* 282:239–40
Walhout AJ, Reboul J, Shtanko O, Bertin N, Vaglio P, et al. 2002. Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr. Biol.* 12:1952–58
Wang X, Grus WE, Zhang J. 2006. Gene losses during human origins. *PLOS Biol.* 4:e52
Whitford WG. 1976. Sweating responses in the chimpanzee (*Pan troglodytes*). *Comp. Biochem. Physiol. A* 53:333–36
Winter H, Langbein L, Krawczak M, Cooper DN, Jave-Suarez LF, et al. 2001. Human type I hair keratin pseudogene phihHaA has functional orthologs in the chimpanzee and gorilla: evidence for recent inactivation of the human gene after the *Pan-Homo* divergence. *Hum. Genet.* 108:37–42
Wood B, Collard M. 1999. Anthropology—the human genus. *Science* 284:65–66
Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–49
Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, et al. 2003. Low nucleotide diversity in chimpanzees and bonobos. *Genetics* 164:1511–18
Yunis JJ, Prakash O. 1982. The origin of man: a chromosomal pictorial legacy. *Science* 215:1525–30
Zhang J. 2003. Evolution of the human ASPM gene, a major determinant of brain size. *Genetics* 165:2063–70
Zhang J, Webb DM, Podlaha O. 2002. Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. *Genetics* 162:1825–35

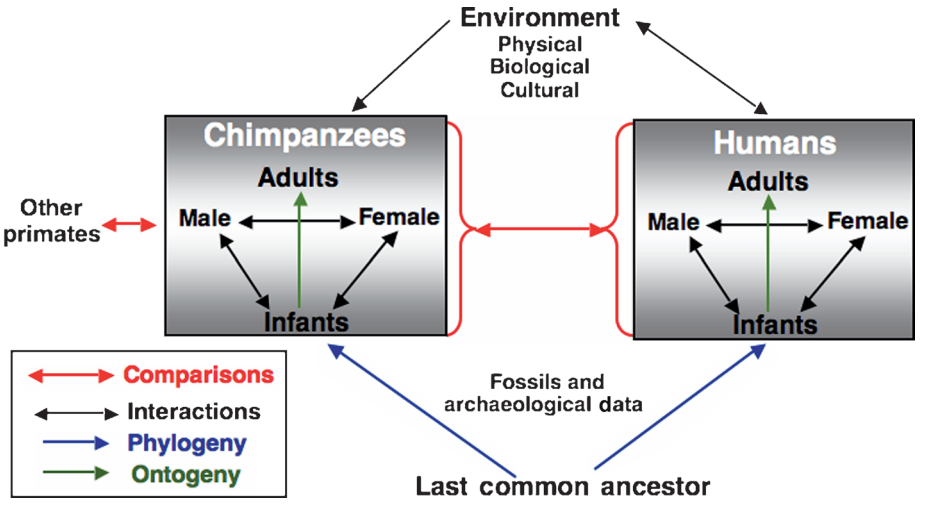

Figure 1

Explaining humans: a multidisciplinary approach. This review is focused on the use of comparative genomics to help explain the origin of humans and unique aspects of our current phenotype. However, any approach to this question must be a multidisciplinary one, which accounts for all relevant information from all areas of human knowledge. This figure attempts to summarize such an approach. Some of the approaches are more important and/or have a stronger support database than do others.

Annual Review of Anthropology
Volume 36, 2007

Contents

Sociocultural Anthropology

Indexes

Errata

An online log of corrections to *Annual Review of Anthropology* articles may be found at http://anthro.annualreviews.org/errata.shtml