

# SEQUENCING THE CHIMPANZEE GENOME: INSIGHTS INTO HUMAN EVOLUTION AND DISEASE

Maynard V. Olson\* and Ajit Varki<sup>†</sup>

Large-scale sequencing of the chimpanzee genome is now imminent. Beyond the inherent fascination of comparing the sequence of the human genome with that of our closest living relative, this project is likely to yield tangible scientific benefits in two areas. First, the discovery of functionally important mutations that are specific to the human lineage offers a new path towards medical benefits. Second, chimpanzee–human comparisons are likely to yield molecular insights into how new biological characteristics evolve — findings that might be relevant throughout the tree of life.

GREAT APES  
Orang-utans, gorillas,  
chimpanzees and bonobos.

\*University of Washington  
Genome Center,  
Departments of Medicine  
and Genome Sciences,  
University of Washington,  
Seattle, Washington 98195,  
USA. <sup>†</sup>Glycobiology Research  
and Training Center,  
Department of Medicine  
and Department of Cellular  
and Molecular Medicine,  
University of California —  
San Diego, La Jolla,  
California 92093-0687,  
USA.  
Correspondence to M.V.O.  
e-mail:  
mvo@u.washington.edu  
doi:10.1038/nrg981

The movement to sequence the genome of the chimpanzee (*Pan troglodytes*) began only five years ago<sup>1</sup>, rapidly gained momentum<sup>2,3</sup> and has now been accorded high priority by the National Human Genome Research Institute of the US National Institutes of Health<sup>4</sup> (BOX 1). The imminent prospect of comparing the chimpanzee and human genome sequences poses new conceptual challenges for comparative genomics, a field that is accustomed to interpreting sequence divergence over much longer timescales than the 5+ million years that have elapsed since the existence of the last common ancestor of chimpanzees and humans. The chimpanzee–human comparison will also focus attention on a central question in evolutionary biology about which we know little: what types of change in what types of gene account for the emergence of a truly new species? It is not just our anthropocentrism that leads us to regard humans as an unusual offshoot of the generally conservative GREAT APES: humans are the conspicuous outliers among primates, and their phenotypic divergence from the great apes was sudden. The challenge will be to find the genetic changes in the human lineage that account for our unusual traits, such as bipedalism and large brains.

Chimpanzee–human comparisons also have more practical implications. Current thinking about the genetic basis of human health and disease is dominated

by studies of differences in disease susceptibility among humans. However, this view might be too parochial. It is likely that the spectrum of common medical ailments from which humans suffer is a characteristic of the species as a whole — a consequence of genetic compromises that occurred during the rapid phenotypic divergence of humans from the great apes. Current knowledge of the biomedical traits that chimpanzees share with humans supports this view. On the one hand, humans and chimpanzees are similar with regard to many aspects of physiology and disease. On the other hand, there are several definite and likely differences (TABLE 1). Here, we focus on these differences, as they are most relevant to the upcoming comparison of the human and chimpanzee genomes. We also emphasize that our knowledge of chimpanzee biology remains incomplete. If we are to make full use of the chimpanzee genome sequence, we must address this imbalance, while respecting the emerging consensus that it is no longer appropriate to use chimpanzees as experimental animals<sup>5</sup>.

In this review, we explore the scientific and medical issues that are relevant to the analysis of the chimpanzee genome sequence. We summarize past speculation about the molecular basis for the differences between chimpanzees and humans and put forward specific hypotheses about the types of functional

## Box 1 | The chimpanzee genome 'project'

The sequencing of the chimpanzee genome is more loosely organized than was the Human Genome Project. The immediate prospect for collecting large-scale data on the whole genome results from the decision of the National Human Genome Research Institute (NHGRI) of the US National Institutes of Health to rate the chimpanzee genome as a high-priority sequencing target<sup>4</sup>. This designation allows large-scale sequencing centres that have NHGRI support to use the sequencing capacity that is released from human and mouse sequencing for the chimpanzee project. The Genome Sequencing Center at Washington University in St Louis, Missouri, and the Whitehead Genome Center at the Massachusetts Institute of Technology have chosen this option and are planning to work together to produce '4×' coverage of the chimpanzee genome by the summer of 2003 (R. H. Waterston, personal communication). At this level of coverage, each base in the genome will have been sampled an average of four times in high-quality data, and the expectation is that >98% of the bases will have been sampled at least once. Long-range assembly of 4× sampling will depend heavily on alignment with the finished human sequence, and there will be many gaps (as discussed later in this review); however, the data will allow most of the sequence differences between the chimpanzee and human genomes to be discovered.

A parallel effort to obtain the complete sequence of chimpanzee chromosome 22, which is homologous to human chromosome 21, has been announced by a Japanese consortium that involves the University of Tokyo Institute of Medical Science and the National Institute of Genetics. This group also aims to complete this sequencing effort during 2003 (REF. 4).

mutation that might be most easily recognized when the chimpanzee and human genomes are compared. Finally, we summarize known differences in the patterns of health and disease between the two species and discuss steps that will need to be taken to maximize the medical benefits of acquiring a chimpanzee genome sequence.

## Molecular comparisons

The biochemical similarities between chimpanzees and humans have been noted for nearly a century. Early evidence was based on immunological studies<sup>6</sup>, which were superseded in the second half of the twentieth century by electrophoretic analyses of blood proteins and by protein sequencing<sup>7–11</sup>. Human–chimpanzee differences in amino-acid sequences vary among proteins but are commonly <1% (REFS 11–13). Direct samplings of chimpanzee genomic DNA indicate that 95% of the chimpanzee genome can be aligned directly with corresponding segments of the human genome<sup>14</sup>, and also indicate that, in these aligned regions, sequence divergence averages 1.2% (REFS 14,15). The 5% that does not align is due to insertions and deletions that have occurred in both chimpanzee and human lineages since their divergence from the last common ancestor. The long-range organization of the two genomes is also highly conserved. The exceptions to this rule are a few discrete differences between the chimpanzee and human karyotypes<sup>16–18</sup>. For example, the chimpanzee has one more chromosome than does the human, because the genetic material on human chromosome 2 is split between chimpanzee chromosomes 12 and 13, and there are also large inversions on human chromosomes 1 and 18 relative to their chimpanzee counterparts. Other smaller rearrangements have been described, and HETEROCHROMATIC genome segments are

also distributed differently on several chromosomes in the two species. Nonetheless, the high-resolution-banding patterns of chimpanzee and human chromosomes are strikingly similar, particularly in EUCHROMATIC portions of the two genomes<sup>16,17</sup>. At the level of biochemical pathways and metabolism, only a few specific differences between great apes and humans are known<sup>13,18,19</sup>.

## Obstacles to studying human evolution

The opportunity to compare the chimpanzee and human genomes is of special importance because we have few other paths towards developing a molecular view of human evolution. In principle, there are three ways to study the molecular basis of recent evolutionary divergences: by analysing clusters of closely related species, by sequencing fossil DNA and by characterizing genetic variation within species. In the case of the human, none of these strategies is likely to have much power because of the lack of closely related sibling species and the limitations on the temporal reach of studies on both fossil DNA and intraspecies variation.

Current views of the phylogenetic relationships between humans and great apes are summarized in FIG. 1 and BOX 2. A key point is the paucity of successful speciation events as a present-day outcome of the past 13+ million years of great ape and human evolution. The fossil record indicates that there were many branches of the human lineage<sup>20</sup>, but that we are the only survivors. Hence, we are unable to put the functionally important mutations that led to modern humans into historical sequence. We must settle instead for inferring the genome sequence of the last common ancestor of chimpanzees and humans and then for examining the cumulative effects of the 5+ million years of evolution along the human lineage.

Although the analysis of fossil DNA from extinct hominid lineages would theoretically allow a stepwise view of human evolution, the practical potential of this approach seems limited. The only well-documented success in analysing fossil DNA from hominids has been the sequencing of portions of the mitochondrial genome from Neanderthal skeletons<sup>21,22</sup>. The best dated of these fossils is only 29,000 years old. The mitochondrial sequences clearly separate the Neanderthals from modern humans and indicate a possible divergence time in the low hundreds of thousands of years. The prospects of extending this type of analysis either to a biologically informative sampling of nuclear genes from Neanderthals or to much older fossils seem poor because of the chemical instability of DNA<sup>23</sup>.

Another path towards understanding human evolution at a molecular level is to study human genetic variation. This approach, too, has limited potential to clarify any but the most recent events. The average nucleotide diversity that is seen among humans (that is, the fraction of base-pair differences expected when comparing two random samplings of the human gene pool) is  $7.5 \times 10^{-4}$  (<1 in 1,000 base pairs)<sup>24</sup>. Theoretical models of the behaviour of genes in populations indicate that this low level of diversity results from a small ancestral population size, which is predicted to have been of the order of

## HETEROCHROMATIN

The densely staining regions of the nucleus that generally contain condensed, transcriptionally inactive regions of the genome.

## EUCHROMATIN

The lightly staining regions of the nucleus that generally contain decondensed, transcriptionally active regions of the genome.

Table 1 | Apparent biomedical differences between humans and great apes\*

Medical condition	Humans	Great apes	References
<b>Definite differences</b>			
HIV progression to AIDS	Common	Very rare	60
Late complications in hepatitis B/C	Frequent	Uncommon	61–64
<i>Plasmodium falciparum</i> malaria	Susceptible	Resistant	65,66
Menopause	Universal	Rare	67
<b>Likely differences</b>			
Influenza A symptomatology	Moderate to severe	Mild	68,69
Alzheimer disease pathology (neurofibrillary tangles)	Common	Rare	70,71
Myocardial infarction	Common	Uncommon	72
Simian foamy virus infection	Very rare	Very common	73–76
Epithelial cancers (carcinomas)	Common	Rare	77,78
<i>Escherichia coli</i> K99 gastroenteritis	Resistant	Sensitive?	79

\*Reviewed in REF. 13. The list excludes conditions explained by obvious anatomical differences between humans and apes (for example, those related to upright posture in humans, or pharyngeal air sacs in great apes). See text for anecdotal examples of other differences. AIDS, acquired immunodeficiency syndrome; HIV, human immunodeficiency virus.

only 10,000 individuals<sup>25</sup>. In populations of this size, most new mutations either become FIXED or, more commonly, are eliminated from the population within a few hundred thousand years. Hence, although studies of human genetic variation might provide insights into genetic processes that were important during the last stages of human evolution (such as the development of complex language and art, and the use of sophisticated tools), they rarely allow us to look more deeply back in time. Most of the history of the 'descent of humans' has been erased from the human gene pool as human-specific alleles became fixed in the whole population.

The paucity of close living relatives of modern humans, the limitations associated with the analysis of fossil DNA and the typically short timescale that is accessible through studies of human genetic variation all highlight the scientific importance of sequencing the chimpanzee genome. Our ability to compare the chimpanzee and human genomes provides us with the best opportunity to look back over millions of years at the genetic events that shaped the emergence of humans as a new offshoot of the great-ape lineage. An initial step in interpreting the chimpanzee sequence will be to test evolutionary hypotheses about the molecular events that underlie rapid evolutionary change.

#### Testing evolutionary hypotheses

Comparison of the chimpanzee and human genomes will break new ground in studies of molecular evolution because it will be the first opportunity to compare the genomes of two species that are only slightly diverged in sequence and yet are on markedly different phenotypic trajectories. We discuss two hypotheses of the genetic basis of the phenotypic differences between chimpanzees and humans that will help guide the comparative interpretation of the genome sequences; one hypothesis has dominated thinking about chimpanzee–human differences for 25 years, whereas the other is more recent.

The classic hypothesis proposes that the crucial genetic differences between chimpanzees and humans involve developmental regulation, whereas the more recent theory proposes that genetic loss on the human lineage has had a crucial role in the emergence of humans. We also discuss recent data that shed light on the general principle that functionally important amino-acid substitutions, as well as gene duplication and divergence, are important mechanisms of evolutionary change.

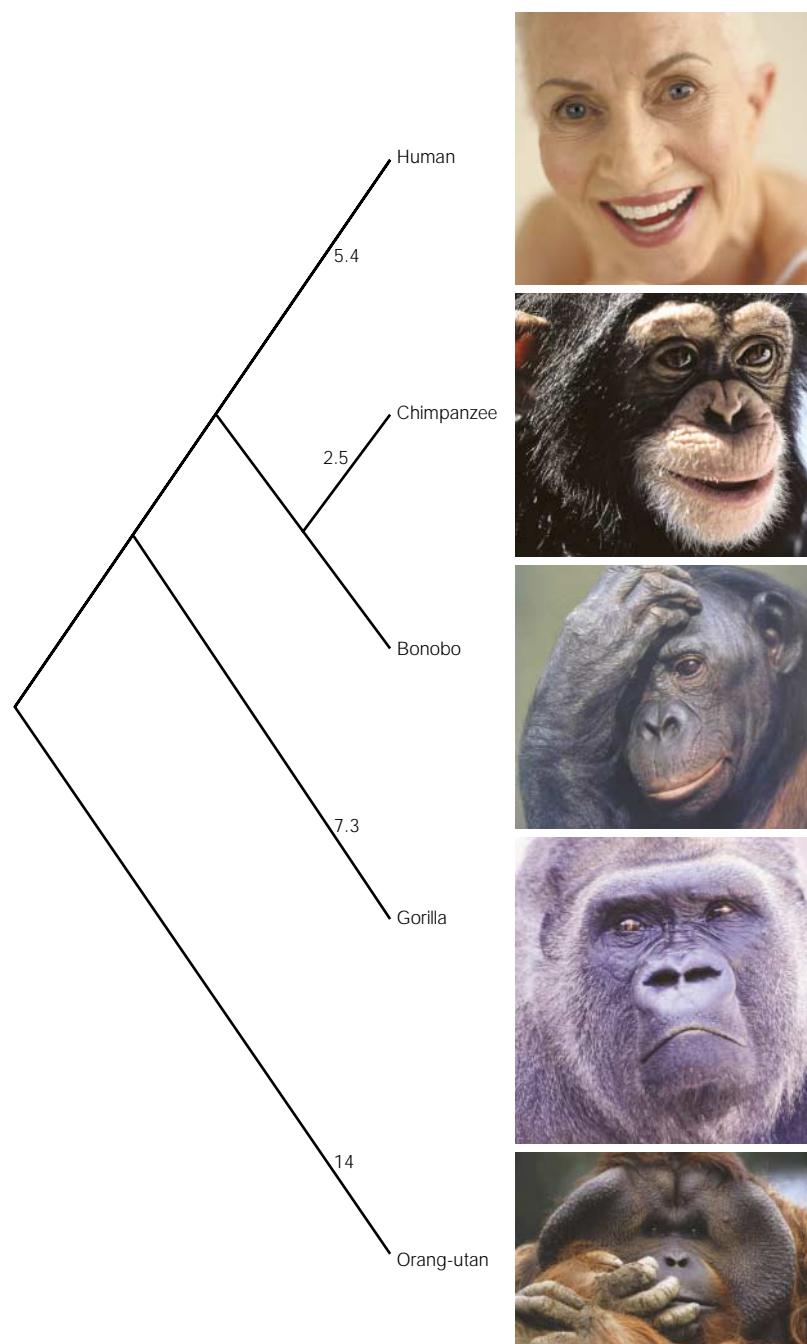
**Differences in gene regulation.** In 1975, Mary-Claire King and Alan Wilson suggested that the phenotypic divergence of chimpanzees and humans is likely to be due primarily to changes in gene regulation rather than to alterations in the protein repertoires of the two species<sup>11</sup>. This idea has its roots in the intuitive notion that slight differences in protein sequence seem inadequate to explain the marked phenotypic differences that are evident between the two species. There are both strengths and weaknesses to this hypothesis. One strength is that many molecular studies of developmental systems, such as the homeotic genes in *Drosophila*, have provided empirical evidence that relatively minor genetic changes can have profound developmental effects when they influence the timing or level of expression of regulatory proteins<sup>26</sup>. However, a weakness of the theory is that we have no real means of estimating how much phenotypic divergence a 1% difference in protein sequence can support. Furthermore, the distinction between sequence alterations in proteins and 'regulatory mutations' has been blurred by the discovery that a significant fraction of all proteins in higher organisms are transcription factors<sup>27</sup>. Because amino-acid substitutions in transcription factors are efficient ways to alter the expression of numerous genes, small differences in protein sequence could themselves cause the developmental effects that King and Wilson sought to explain.

#### FIXATION

The process whereby a genetic variant arises by mutation and then increases in frequency in a population until it is the only variant present.

#### TRANSTHYRETIN

A protein that is a key transporter of the thyroid hormones in the blood and cerebrospinal fluid. Thyroid hormones stimulate an increase in the metabolic rate of many cell types and have effects on embryonic brain development.



**Figure 1 | Sequence-based phylogenetic tree for the human and the great apes.**

The numbers indicate estimated millions of years (Myr) from the last branch point to the present along particular lineages. The tree was largely redrawn from data in REF. 48. The time estimates are based on the assumption that humans and orang-utans diverged 14 Myr ago and that MOLECULAR-CLOCK rates are equal on all lineages. The chimpanzee–bonobo split has not been well calibrated with sequence data: the estimate shown is based on the fragmentary data in REF. 59. Bonobo photograph by Frans Lanting, reproduced with permission from REF. 54 © Frans Lanting.

#### MOLECULAR CLOCK

The steady accumulation of mutations during evolution, which provides a basis for dating the point at which two contemporary species diverged from a common ancestor.

The sequencing of the chimpanzee genome will greatly facilitate experimental tests of specific genetic differences between chimpanzees and humans that seem likely to influence developmental regulation. Experimental tests could be carried out *in vitro*, in transfected cells or in transgenic animal models such as

the mouse. Given that tests in transgenic animals will be expensive — and will largely require extrapolation to the chimpanzee and human from results on distantly related organisms — it will be important to develop *in vitro* and cell-culture assays that are as informative as possible about the whole-organism biology of the two species. When chimpanzee–human differences involve genetic characteristics that chimpanzees share with a wide range of other primates, whole-animal experiments on the rhesus macaque, or on other primates that are commonly used in biomedical research, will provide an important means of validating the functional significance of the differences.

To some degree, it might be possible to correlate observed changes in genome sequence with differences in gene-expression patterns between chimpanzee and human tissues. Studies have already begun on expression-pattern differences<sup>28</sup>, and they will undoubtedly accelerate as the chimpanzee and human genome sequences allow progressively more sophisticated expression arrays to be designed. A recent study of mRNA levels in various chimpanzee and human tissues indicates that there are more marked differences in expression patterns in tissues that are derived from the central nervous system, particularly the cerebral cortex, than for example the liver<sup>28</sup>. These data indicate that the evolutionary rate of change of gene expression in the human brain might have been faster than that found in organs with fewer biological differences between apes and humans, such as the liver. Patterns of liver and blood-cell gene expression in chimpanzees and humans seemed similar, whereas those in the rhesus monkey liver seemed less similar. This result is as expected on the basis of the relative divergence times of the three lineages. By contrast, patterns of gene expression in the cerebral cortex in chimpanzees and monkeys were more similar, whereas humans were the outliers. If validated by further studies, these results offer the hope that gene-expression studies might explain the marked changes in cognitive function that accompanied human evolution. However, the potential for artefacts in such studies — given their reliance on autopsy material from both species and on the anatomical complexity of primate brains — is sobering. For example, it is possible that there are age or species-specific differences in the rate at which particular brain tissues deteriorate after death — or simply differences in the handling of chimpanzee and human tissue that are difficult to eliminate.

Despite the experimental difficulties of designing tests of the King–Wilson hypothesis, it is likely that some portion of the phenotypic differences between humans and great apes is due to differences in the timing, level and pattern of gene expression. One example that seems to fit this model is the lower level of TRANSTHYRETIN expression in humans compared with chimpanzees<sup>29</sup>. The qualitative change in the amount of this protein seems to correlate with altered thyroid hormone metabolism, which is likely to influence a wide variety of phenotypes, such as brain development and function, skull morphology and metabolic rate.

## Box 2 | The evolution of humans and great apes

DNA-sequence data strongly support a phylogenetic tree for humans and the great apes in which the order of divergence from the human lineage was: orang-utan, gorilla, chimpanzee<sup>48,52</sup> (FIG. 1). Although there is molecular evidence for the existence of sub-species of chimpanzees, gorillas and orang-utans<sup>53</sup>, only the chimpanzee lineage has split convincingly into two surviving species, the common chimpanzee (*Pan troglodytes*) and the bonobo (*Pan paniscus*)<sup>54</sup>. Although the tree in FIG. 1 provides a reliable view of the relative timing of key events in the evolution of humans and great apes, its absolute calibration depends on molecular-clock extrapolations from much deeper branches of the mammalian evolutionary tree. Only over a time span of tens of millions of years is the fossil record sufficiently rich to allow reliable correlations to be made between evolutionary events and geological dates. Extrapolations from these data indicate that humans and chimpanzees probably diverged ~5.5 million years ago<sup>48,55,56</sup>. However, this date remains uncertain and seems more likely to be revised upwards than downwards; for example, one recently described hominid fossil that is judged to post-date the human–chimpanzee divergence seems to be 6–7 million years old<sup>57</sup>.

## PURIFYING SELECTION

The elimination of deleterious mutations through natural selection.

## SIALIC ACID

An acidic sugar that is commonly found at the ends of the glycan chains of cell-surface glycoproteins and glycolipids. They are negatively charged under physiological conditions, contribute to biophysical characteristics of cell surfaces and can be recognized by many receptors of endogenous and exogenous origin.

## ALU

A dispersed, intermediately repetitive, 300-bp DNA sequence present in the human genome.

## FRAMESHIFT MUTATION

A mutation that results in a change in the reading frame of a protein-encoding region. Frameshift mutations frequently cause such marked changes in a protein sequence that the protein is completely inactivated.

## NONSENSE MUTATION

A mutation that results in the introduction of a stop codon to cause the premature termination of a protein. Nonsense mutations often completely inactivate a protein.

## NON-SYNONYMOUS SUBSTITUTION

A mutation in the coding region of a gene that changes the amino acid inserted at a particular position in a protein.

**The 'less-is-more' hypothesis.** An alternative to the King–Wilson hypothesis is Olson's 'less-is-more' hypothesis, which emphasizes the importance of loss-of-function mutations on a recently evolved novel lineage such as the human<sup>30</sup>. This hypothesis, which encompasses the effects of mutations in genes that encode regulatory, catalytic and structural proteins, is counter-intuitive as it adopts the premise that the human, in many biological respects, is a 'degenerate ape'. To some degree, genetic loss in the human lineage seems almost certain to have caused some of the differences that exist between chimpanzees and humans: examples such as delayed postnatal development and loss of muscle strength and hair in humans all seem 'degenerative'. However, it is unclear how far it is safe to extrapolate from these conspicuously retrograde phenotypes to broader aspects of human and chimpanzee biology.

The great advantage of genetic loss as a mechanism for phenotypic evolution is its potential to occur rapidly. Indeed, if, as seems likely, PURIFYING SELECTION is relaxed for a species that has suddenly broken out from the selective constraints that have long applied to a stable taxonomic group, a cascade of genetic loss is inevitable. Although genetic loss is unlikely to account for the initial 'breaking out' of a new lineage, the genetic alterations that lead to the launch of such a lineage are likely to be a small proportion of the genetic differences that contribute to its subsequent phenotypic 'makeover'. Therefore, even if new gene functions have a crucial causal role in the evolution of a new species, subsequent genetic-loss events might greatly outnumber the early genetic innovations.

Existing data bearing on the less-is-more hypothesis are tantalizing but limited. Although the best-described differences in gene function between chimpanzees and humans all support the hypothesis, they are few in number. The first clear example of a major biochemical difference between chimpanzees and humans was the discovery that, unlike great apes, humans cannot synthesize a form of the cell-surface SIALIC ACID called *N*-glycolylneuraminic acid (Neu5Gc)<sup>31</sup>. A secondary consequence is an excess of its precursor *N*-acetylneuraminic acid (Neu5Ac). The synthesis of Neu5Gc, which is widely

expressed on cell surfaces in many tissues in non-human primates, depends on the hydroxylation of CMP-Neu5Ac. In all humans, the CMP-Neu5Ac hydroxylase (*CMAH*) gene has been inactivated by a 92-bp deletion that occurred in the human lineage after the divergence of humans and chimpanzees<sup>32,33</sup>. This deletion was the by-product of the insertion of a new human-specific *ALU* sequence into the hydroxylase gene, which replaced an ancient *Alu* element<sup>34</sup>. Studies of sialic acids from Neanderthal fossils, molecular-clock analysis of the human *CMAH* gene and comparisons of the new *Alu* with other members of its sub-family indicate that the mutation might have occurred ~2.5–3 million years ago<sup>35</sup>. The changes that are associated with bipedalism considerably pre-date this mutation. However, the apparent date of the mutation indicates that the inactivation of *CMAH* probably occurred before the onset of brain expansion<sup>20</sup>. Knowledge of the biological functions of this specific sialic acid is, as yet, insufficient to relate this change to particular human-specific characteristics<sup>36</sup>.

These alterations in sialic acid biochemistry are of special interest as they involve genetic loss in a metabolic system that lacks extensive redundancy. As such, the mutations are likely to have direct biochemical consequences. Several other examples of genetic loss in the human lineage are known, but they all involve the loss of one member of a large gene family, the members of which are likely to share overlapping functions; they include the *V10* variable gene of the human T-cell-receptor- $\gamma$  locus<sup>37</sup>, the olfactory receptor gene *OR 912–93* (REF. 38) and a type I hair-keratin gene<sup>39</sup>.

A strength of the less-is-more hypothesis, in addition to its genetic plausibility, is that it is readily testable. The testability of the hypothesis rests on the relative ease with which loss-of-function mutations can be recognized simply by sequence analysis. Many human genetic diseases, such as **cystic fibrosis**, **phenylketonuria** and the forms of familial **breast cancer** caused by mutations in the breast cancer 1, early onset (*BRCA1*) gene and *BRCA2*, seem simply to require loss of the relevant gene function. For these diseases, thousands of independent mutations have been detected in patients who have been identified on the basis of disease phenotype and family history. Presumably, these mutations are a reasonable sample of the ways in which gene loss occurs in humans. In general, the mutations are easily recognized: FRAMESHIFT MUTATIONS, NONSENSE MUTATIONS and splice-site alterations account for a substantial majority of loss-of-function mutations that cause monogenic diseases<sup>40</sup>. Of course, there are some NON-SYNONYMOUS SUBSTITUTIONS that cannot be easily distinguished from neutral polymorphisms, but the general pattern is principally one in which mutations have obviously deleterious effects on gene function. Hence, the less-is-more hypothesis makes the readily testable prediction that the initial comparison of the chimpanzee and human genomes should show a substantial number of conspicuous loss-of-function mutations in the human lineage. If true, the list of these genes will provide a powerful starting point from which to explore the many differences between chimpanzee and human biology.

**Other genetic mechanisms of divergence.** Traditional ideas about the genetic changes that lead to the evolution of new biological characteristics emphasize gene duplication and divergence, as well as non-synonymous substitutions in single-copy genes. Recent analyses of the reference sequence of the human genome indicate that gene duplication and divergence can lead to rapid, large-scale alterations in the genetic repertoire of a species. Approximately 5% of the euchromatic portion of the human genome consists of segmental duplications of >1 kb with >90% sequence similarity<sup>41,42</sup>. This level of sequence similarity is comparable with the similarity of human genes and their homologues in OLD WORLD MONKEYS. Hence, duplications that meet the criterion of >90% sequence similarity between the duplicated copies could be as old as 30 million years. However, there is some indication that duplication of genome segments in the primate lineage has been most active during the past 15 million years<sup>41</sup>. Furthermore, the process is clearly continuing: instances of small-scale duplications that post-date the human–chimpanzee divergence are already known<sup>43</sup>, and such duplications could prove to be relatively abundant once whole-genome comparisons are possible. Detailed analyses of recent duplications in the human genome indicate that rapid divergence of duplicated genes sometimes occurs<sup>41</sup>. One example of a gene duplication that is specific to humans is the gene that encodes protocadherin XY, which lies in a region of the human Y chromosome that was duplicated and translocated from the X chromosome after the last common ancestor of chimpanzees and humans. It has been postulated that protocadherin XY is involved in handedness, language lateralization and brain asymmetry<sup>44,45</sup>. Hence, although the analysis of recent duplications that are specific to the human lineage has just begun, early results support the possibility that gene duplication and divergence have been important contributors to the phenotypic divergence of humans and chimpanzees.

Another class of potentially relevant mutations are those that change protein function by altering amino acids, particularly amino acids that are typically evolutionarily conserved. For example, a single-nucleotide change led to an amino-acid substitution in human **SIGLECL1**, which eliminates the ability of the protein to interact efficiently with Neu5Gc and other sialic acids<sup>46</sup>. Another recent and exciting example involves two amino-acid changes in the highly conserved forkhead transcription factor **FOXP2**, a protein that has been implicated through genetic studies in human speech and language. The amino-acid changes and the pattern of nucleotide polymorphism in human **FOXP2** strongly indicate that it has been the target of selection during recent human evolution<sup>47</sup>.

Regardless of the type of mutation that is a candidate for contributing to differences between chimpanzees and humans — such as regulatory alterations, genetic loss, functional amino-acid changes and gene duplication and divergence — it is essential to know whether the mutation occurred on the chimpanzee or

the human lineage. Additional data are required, beyond the chimpanzee and human genome sequences, to assign mutations to the appropriate lineage. In most cases, it will be possible to infer correctly which variant was present in the last common ancestor of chimpanzees and humans simply by comparing the two sequences with a sequence from the gorilla or the orang-utan. Because sequence divergence among all the great apes is small, most mutations that have occurred since the last common ancestor of humans and great apes would have only occurred on one of the lineages. The other species would share a sequence that is likely to have also been present in the last common ancestor of humans and great apes. The gorilla would be the best choice for a comparison species because its genome sequence resembles that of the last common ancestor between chimpanzees and humans more closely than does that of the orang-utan<sup>48</sup>. A reasonable approach would be to carry out whole-genome sampling of the gorilla to a modest level of coverage, such as 3× (that is, a level of coverage that would sample single-copy sequences in the gorilla an average of three times). At 3× sampling, there would be 95% coverage of the gorilla genome, and ancestral states would be resolved, with relatively few errors, for 95% of the ~40-million (~1.2% of  $3 \times 10^9$  bp) chimpanzee–human discrepancies. An early move to sequence the genome of the rhesus macaque would also greatly enhance the reliability of ancestral-state inferences by adding data from a fourth closely related species to the analysis.

Finally, it should be pointed out that the ability to recognize functionally important mutations against a background of tens of millions of neutral changes, and to infer ancestral states, will depend greatly on the quality of the available sequence data. Although a high-quality human sequence is now assured, the quality goals for chimpanzee sequencing have still to be resolved, as it is not yet clear if finished chimpanzee sequence will be produced. The cost–benefit issues surrounding this issue are discussed in BOX 3.

#### Biomedical significance

Humans and chimpanzees are similar with regard to many aspects of physiology and disease. Here, we focus primarily on known and apparent differences (TABLE 1), because these are most relevant to the upcoming comparison of the human and chimpanzee genomes. The chimpanzee genome project offers a unique opportunity to step back and look at the disease susceptibilities of our species. Humans that live under modern conditions have a characteristic set of susceptibilities: numerous infectious diseases that are human specific (or nearly so); cardiovascular disease; **CARCINOMAS**; obesity; **type II diabetes**; autoimmune diseases; major psychoses; and neurodegenerative diseases. Although individual humans vary in their genetic susceptibilities to these conditions, we should not allow these intra-species differences to distract attention from the larger question of why humans have this particular spectrum of susceptibilities. To some extent, the unique environments in which humans live

OLD WORLD MONKEYS  
Monkeys that are native to Africa and Asia.

CARCINOMA  
A type of cancer that originates from epithelial cells. Most human cancers other than leukaemias or lymphomas are carcinomas.

**HYDATIDIFORM MOLAR PREGNANCY**

A pregnancy resulting from an abnormal fertilization event whose product can expand through successive cell divisions but cannot undergo normal development.

**ELECTROPHEROGRAMS**

Raw data that are produced during DNA sequencing. An electropherogram displays the fluorescence produced by DNA molecules that have been electrophoretically separated during the DNA sequencing process.

undoubtedly contribute to human patterns of health and disease. However, the environments in which captive chimpanzees live increasingly resemble those of humans. Hence, genetic differences are likely to contribute to differences in the typical health profiles of humans and captive chimpanzees, and the identification of the crucial differences would be of great biomedical interest.

Our existing knowledge of the disease profiles of chimpanzees is fragmentary but intriguing. TABLE 1 summarizes examples of chimpanzee–human differences that are either well established or indicated by existing reports<sup>13</sup>. In addition to the conditions listed in TABLE 1, anecdotal reports indicate that some common human conditions are rare in chimpanzees. These include early fetal wastage, HYDATIDIFORM MOLAR PREGNANCY, bronchial asthma, acne vulgaris, major psychoses, and autoimmune diseases such as **systemic lupus**, **rheumatoid arthritis** and **multiple sclerosis** (K. Benirschke and B. Swenson, personal communications; A.V., unpublished observations). A striking feature of this list is that it includes many of the common diseases of central importance to contemporary biomedical research in economically developed countries.

A programme to understand the genetic basis of human susceptibility to these diseases would open new avenues to the development of improved therapies. Perhaps during the 'hasty makeover' through which a great ape ancestor evolved into modern humans, specific compromises in human physiology made us susceptible to the diseases that are now the predominant causes of human morbidity and mortality. Although development of the comparative medicine of chimpanzees and humans into a principal area of investigation would

represent a high-risk–high-gain strategy in biomedical research, the potential pay-off justifies the risk that the differences between these sibling species will prove too difficult to interpret.

If the less-is-more hypothesis actually accounts for a significant proportion of the most important physiological differences between chimpanzees and humans, the task of developing a molecular analysis of the crucial genetic differences might prove less formidable than it first seems. In this case, the human-specific mutations of large functional effect will be relatively easy to recognize. Furthermore, we would have the example of other primates in which to study how animals whose biology is generally similar to ours avoid the characteristic human pathologies. The medical goal would be to learn how to emulate aspects of primate biology that have been lost in humans. This concept has the potential to open a new front in biomedical research that differs fundamentally from current models.

**Ethical considerations**

Bioinformatic analysis of the chimpanzee and human reference sequences will only take us so far. Greatly expanded studies of the veterinary medicine and basic biology of captive chimpanzees will also be essential. Chimpanzees breed well in captivity, to the point that rigorous birth-control programmes are in effect in most colonies<sup>50</sup>. However, the future of these colonies poses substantial ethical, logistical and policy challenges; among these is the need to develop a coherent view on the ethics of experimentation on chimpanzees.

The chimpanzee is no longer regarded as a conventional experimental animal<sup>50</sup>. Indeed, we believe that there is an emerging social consensus that humans have an ethical commitment to preserve the remaining great ape populations in the wild and to maintain captive populations under conditions that maximize the quality of life of their members. The acquisition of a chimpanzee genome sequence, and the attendant increased focus on chimpanzee research, should be accompanied by careful study of how humans can best meet our stewardship responsibilities towards these close kin. A likely conclusion is that the standards for approving proposals to carry out research on captive chimpanzees should resemble those that apply to humans who are not considered able to provide informed consent. These standards, which have received considerable attention<sup>51</sup>, include the principle that research risks to the individual must be small and offset by potential benefits. Although these broad principles provide a starting point for developing a social consensus about the appropriate future use of chimpanzees in biomedical research, the means of implementing them will require careful study. Just as the Human Genome Project was coupled from its inception with a programme to study the ethical, legal and social implications of genome research, we suggest that comparable attention to these issues, as they apply to research on chimpanzees, should be an integral part of the chimpanzee genome project. We expect the proper treatment of captive chimpanzees to be of intense interest to scientists and non-scientists alike.

**Box 3 | The case for a high-quality chimpanzee sequence**

In current practice, genome sequencing occurs in two steps. First, most of the data are acquired by random sampling of the sequencing target in short sequencing 'reads'. These individual ELECTROPHEROGRAMS typically contain ~500 nucleotides of high-quality data. In the case of the chimpanzee project, most of the data will undoubtedly be acquired simply by whole-genome sampling of chimpanzee DNA rather than by the 'clone-by-clone' method<sup>58</sup> on which the production of a finished human sequence has relied. Whole-genome sampling is favoured for the chimpanzee because, given the availability of a high-quality human sequence, the assembly of whole-genome shotgun reads of chimpanzee genomic sequence can proceed by aligning these reads with finished human sequence. Second, in the production of a high-quality sequence, is a 'finishing' process during which targeted methods are used to improve data quality in places where sequence is absent or likely to contain errors. To minimize the cost of the project, there will be inevitable pressure to leave the chimpanzee sequence at a 'rough-draft', rather than at a finished, state.

Although a rough-draft sequence would allow most chimpanzee–human differences to be discovered, there would be significant drawbacks to this approach. Sequence quality deteriorates at the ends of contiguous blocks of sequence; hence, each of the gaps in a rough-draft sequence would be a potential site of missed or erroneously interpreted chimpanzee–human differences. More fundamentally, comparison of a rough-draft chimpanzee sequence, largely assembled by alignment with a finished human sequence, risks a serious bias against rapidly evolving segments of the genome. Regions that have undergone duplication on one lineage or the other since the last common ancestor, as well as those that have been selectively deleted on the human lineage — all of which are prime candidates for explaining the phenotypic divergence between the two species — would be particularly vulnerable to misassembly.

## Need for concomitant phenotyping studies

It is essential that the sequencing of the chimpanzee genome be combined with more detailed studies of the normal physiology, pathology and overall biology of the great apes. In these domains, we know far less about great apes than we know about humans. Towards this end, a 'Great Ape Phenome Project' has been proposed to make optimum use of the chimpanzee genome sequence<sup>49</sup>. Such a project should include gathering information and tissue samples during the routine veterinary care of captive great apes and supplementing these resources with other sources of data and research materials that can be gathered without risk to these animals. For example, thorough autopsies with standardized data capture should become common practice at zoos and captive great ape colonies. This practice will teach us much about chimpanzee pathology and normal ageing, and will enable tissue banks and tissue-specific cDNA libraries to be established. These materials will be important resources for the functional annotation of the chimpanzee genome sequence. Phenotypic studies of captive chimpanzees will largely depend on the same medical procedures that are routinely applied to humans. Clinical research on the appropriate application of these procedures to chimpanzees will yield useful data and will also benefit chimpanzees, as standards of care based on human medical knowledge are not always appropriate for these animals.

## Conclusions

The idea of investing heavily in the comparative biology of chimpanzees and humans runs counter to the way

most molecular biologists think about evolutionary questions. A common perspective is that there are no significant differences between chimpanzees and humans, at least in the types of basic cellular process that dominate current research in cell and molecular biology. Efforts to understand the evolution of life are sometimes compared to the peeling of an onion, with the surface layer corresponding to the current biosphere. Molecular biologists like to focus on the deep layers of the onion. In these layers, and in the onion's core, the important molecular processes that characterize all contemporary species became established.

The idea of carrying out a detailed comparison of chimpanzees and humans has different conceptual origins and goals. In this case, the plan is to peel away one layer of the onion in a single taxonomic group. In so doing, we have the potential to reconstruct with minimal ambiguity the genome sequence of the last common ancestor of chimpanzees and humans. Then, in exploring the genetic differences that have become fixed on both lineages, we have the potential to define, in fine-grained detail, the separate biological trajectories that have produced these two species that are so alike and yet so different. The challenge of distinguishing the phenotypically important signals from the neutral noise will be enormous. However, we already have some hypotheses to guide this process and can expect more ideas and experimental options to emerge as biology engages the grand challenge that chimpanzee-human comparisons will pose. Success in this venture has the potential to transform the way we think about health and disease in humans and great apes, as well as human origins and the origins of other novel evolutionary lineages.

- McConkey, E. H. & Goodman, M. A human genome evolution project is needed. *Trends Genet.* **13**, 350–351 (1997).
- McConkey, E. H. & Varki, A. A primate genome project deserves high priority. *Science* **289**, 1295–1296 (2000).
- Normile, D. Genomics: chimp sequencing crawls forward. *Science* **291**, 2297 (2001).
- Cyranoski, D. Almost human. *Nature* **418**, 910–912 (2002).
- Goodman, S. & Check, F. The great primate debate. *Nature* **417**, 684–687 (2002).
- Nutall, G. H. F. *Blood Immunity and Blood Relationships* (Cambridge Univ. Press, London, 1904).
- Washburn, S. L. (ed.) *Classification and Human Evolution* (Aldine, Chicago, Illinois, 2002).
- Syner, F. N. & Goodman, M. Differences in the lactic dehydrogenases of primate brains. *Nature* **209**, 426–428 (1966).
- Sarich, V. M. & Wilson, A. C. Immunological time scale for hominid evolution. *Science* **158**, 1200–1203 (1967).
- Doolittle, R. F., Wooding, G. L., Lin, Y. & Riley, M. Hominid evolution as judged by fibrinopeptide structures. *J. Mol. Evol.* **1**, 74–83 (1971).
- King, M.-C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).  
**Classic paper that reviews earlier literature on the biochemical similarities between chimpanzee and human, adds new data based on serum proteins and proposes that regulatory changes probably account for the rapid phenotypic divergence of chimpanzees and humans.**
- Eyre-Walker, A. & Keightley, P. D. High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347 (1999).
- Varki, A. A chimpanzee genome project is a biomedical imperative. *Genome Res.* **10**, 1065–1070 (2000).  
**Survey of existing knowledge about the biomedical differences between great apes and humans.**
- Britten, R. J. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl Acad. Sci. USA* **99**, 13633–13635 (2002).
- Fujiyama, A. *et al.* Construction and analysis of a human-chimpanzee comparative clone map. *Science* **295**, 131–134 (2002).  
**Analysis of the first large set of genomic DNA sequences from the chimpanzee.**
- Yunis, J. J., Sawyer, J. R. & Dunham, K. The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. *Science* **208**, 1145–1148 (1980).
- Yunis, J. J. & Prakash, O. The origin of man: a chromosomal pictorial legacy. *Science* **215**, 1525–1530 (1982).
- Gagneux, P. & Varki, A. Genetic differences between humans and great apes. *Mol. Phylogenet. Evol.* **18**, 2–13 (2001).
- Hacia, J. G. Genome of the apes. *Trends Genet.* **17**, 637–645 (2001).
- Wood, B. & Collard, M. Anthropology — the human genus. *Science* **284**, 65–71 (1999).
- Krings, M. *et al.* Neandertal DNA sequences and the origin of modern humans. *Cell* **90**, 19–30 (1997).
- Ovchinnikov, I. V. *et al.* Molecular analysis of Neandertal DNA from the northern Caucasus. *Nature* **404**, 490–493 (2000).
- Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M. & Pääbo, S. Ancient DNA. *Nature Rev. Genet.* **2**, 353–359 (2001).
- International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **15**, 928–933 (2001).
- Harpending, H. C. *et al.* Genetic traces of ancient demography. *Proc. Natl Acad. Sci. USA* **95**, 1961–1967 (1998).
- Lewis, E. B. The bithorax complex: the first fifty years. *Int. J. Dev. Biol.* **42**, 403–415 (1998).
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Enard, W. *et al.* Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).
- Gagneux, P. *et al.* Proteomic comparison of human and great ape blood plasma reveals conserved glycosylation and differences in thyroid hormone metabolism. *Am. J. Phys. Anthropol.* **115**, 99–109 (2001).
- Olson, M. V. When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* **64**, 18–23 (1999).  
**Short review outlining the argument that gene loss is likely to be an important mechanism of rapid evolutionary change.**
- Muchmore, E. A., Diaz, S. & Varki, A. A structural difference between the cell surfaces of humans and the great apes. *Am. J. Phys. Anthropol.* **107**, 187–198 (1998).
- Chou, H. H. *et al.* A mutation in human CMP-sialic acid hydroxylase occurred after the *Homo-Pan* divergence. *Proc. Natl Acad. Sci. USA* **95**, 11751–11756 (1998).
- Irie, A., Koyama, S., Kozutsumi, Y., Kawasaki, T. & Suzuki, A. The molecular basis for the absence of *N*-glycolylneuraminic acid in humans. *J. Biol. Chem.* **273**, 15866–15871 (1998).
- Hayakawa, T., Satta, Y., Gagneux, P., Varki, A. & Takahata, N. Alu-mediated inactivation of the human CMP-*N*-acetylneuraminic acid hydroxylase gene. *Proc. Natl Acad. Sci. USA* **98**, 11399–11404 (2001).
- Chou, H. H. *et al.* Inactivation of CMP-*N*-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl Acad. Sci. USA* **99**, 11736–11741 (2002).  
**Reviews the discovery that humans are unable to synthesize a sialic acid found in great apes and other primates due to an inactivating mutation in the gene for an enzyme that modifies a precursor; also, estimates the timing of this mutation relative to other steps in human evolution.**
- Varki, A. Loss of *N*-glycolylneuraminic acid in humans: mechanisms, consequences and implications for hominid evolution. *Yb. Phys. Anthropol.* **44**, 54–69 (2002).

37. Zhang, X. M., Cathala, G., Souza, Z., Lefranc, M. P. & Huck, S. The human T-cell receptor  $\gamma$  variable pseudogene V10 is a distinctive marker of human speciation. *Immunogenetics* **43**, 196–203 (1996).
38. Rouquier, S. *et al.* A gene recently inactivated in human defines a new olfactory receptor family in mammals. *Hum. Mol. Genet.* **7**, 1337–1345 (1998).
39. Winter, H. *et al.* 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 (2001).
40. Jinnah, H. A., De Gregorio, L., Harris, J. C., Nyhan, W. L. & O'Neill, J. P. The spectrum of inherited mutations causing HPRT deficiency: 75 new cases and a review of 196 previously reported cases. *Mutat. Res.* **463**, 309–326 (2000).
41. Samonte, R. V. & Eichler, E. E. Segmental duplications and the evolution of the primate genome. *Nature Rev. Genet.* **3**, 65–72 (2002).
42. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
43. Trask, B. J. *et al.* Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Mol. Genet.* **7**, 13–26 (1998).
44. Crow, T. J. Handedness, language lateralisation and anatomical asymmetry: relevance of protocadherin XY to hominid speciation and the aetiology of psychosis. *Br. J. Psychiatry* **181**, 295–297 (2002).
45. Nicholson, T. R., Yang, J., DeLisi, L. E. & Crow, T. J. Allele sharing for schizophrenia and schizo-affective disorder within a region of *Homo sapiens* specific XY homology. *Am. J. Med. Genet.* **114**, 637–640 (2002).
46. Angata, T., Varki, N. M. & Varki, A. A second uniquely human mutation affecting sialic acid biology. *J. Biol. Chem.* **276**, 40282–40287 (2001).
47. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
48. Chen, F. C. & Li, W. H. 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**, 444–456 (2001).
49. Varki, A. *et al.* Great ape phenome project? *Science* **282**, 239–240 (1998).
50. Committee on Long-Term Care of Chimpanzees, National Research Council. *Chimpanzees in Research: Strategies for their Ethical Care, Management, and Use* (National Academy Press, Washington, DC, 1997).
51. Knoppers, B. M., Avar, D., Cardinal, G. & Glass, K. C. Children and incompetent adults in genetic research: consent and safeguards. *Nature Rev. Genet.* **3**, 221–225 (2002).
52. Goodman, M. The genomic record of humankind's evolutionary roots. *Am. J. Hum. Genet.* **64**, 31–39 (1999).
53. Gagneux, P. *et al.* Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc. Natl Acad. Sci. USA* **96**, 5077–5082 (1999).
54. de Waal, F. & Lanting, F. *Bonobo: The Forgotten Ape* (Univ. of California Press, Berkeley, California, 1997).
55. Takahata, N. & Satta, Y. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proc. Natl Acad. Sci. USA* **94**, 4811–4815 (1997).
56. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
57. Brunet, M. *et al.* A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**, 145–151 (2002).
58. Olson, M. V. The maps: clone by clone by clone. *Nature* **409**, 816–818 (2001).
59. Bailey, W. J. *et al.* Re-examination of the African hominoid trichotomy with additional sequences from the primate  $\beta$ -globin gene cluster. *Mol. Phylogenet. Evol.* **1**, 97–135 (1992).
60. Novembre, F. J. *et al.* Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. *J. Virol.* **71**, 4086–4091 (1997).
61. Krawczynski, K., Prince, A. M. & Nowoslawski, A. Immunopathologic aspects of the HBsAg carrier state in chimpanzees. *J. Med. Primatol.* **8**, 222–232 (1979).
62. Shouval, D. *et al.* Chronic hepatitis in chimpanzee carriers of hepatitis B virus: morphologic, immunologic, and viral DNA studies. *Proc. Natl Acad. Sci. USA* **77**, 6147–6151 (1980).
63. Dienes, H. P., Purcell, R. H., Popper, H. & Ponzetto, A. The significance of infections with two types of viral hepatitis demonstrated by histologic features in chimpanzees. *J. Hepatol.* **10**, 77–84 (1990).
64. Gagneux, P. & Muchmore, E. A. in *HBV Protocols* (ed. Hamatake, R.) (Humana Press, Totowa, New Jersey, in the press).
65. Escalante, A. A., Barrio, E. & Ayala, F. J. Evolutionary origin of human and primate malarial: evidence from the circumsporozoite protein gene. *Mol. Biol. Evol.* **12**, 616–626 (1995).
66. Ollomo, B. *et al.* Lack of malaria parasite transmission between apes and humans in Gabon. *Am. J. Trop. Med. Hyg.* **56**, 440–445 (1997).
67. Graham, C. E. Reproductive function in aged female chimpanzees. *Am. J. Phys. Anthropol.* **50**, 291–300 (1979).
68. Murphy, B. R. *et al.* in *Chimpanzee Conservation and Public Health: Environments for the Future* 21–27 (Diagnon/Bioqual, Inc., Rockville, Maryland, 1992).
69. Subbarao, K., Webster, R. G., Kawaoka, Y. & Murphy, B. R. Are there alternative avian influenza viruses for generation of stable attenuated avian-human influenza A reassortant viruses? *Virus Res.* **39**, 105–118 (1995).
70. Gearing, M., Rebeck, G. W., Hyman, B. T., Tigges, J. & Mirra, S. S. Neuropathology and apolipoprotein E profile of aged chimpanzees: implications for Alzheimer disease. *Proc. Natl Acad. Sci. USA* **91**, 9382–9386 (1994).
71. Gearing, M., Tigges, J., Mori, H. & Mirra, S. S.  $\beta$ -Amyloid (A $\beta$ ) deposition in the brains of aged orangutans. *Neurobiol. Aging* **18**, 139–146 (1997).
72. Schmidt, R. E. Systemic pathology of chimpanzees. *J. Med. Primatol.* **7**, 274–318 (1978).
73. Bieniasz, P. D. *et al.* A comparative study of higher primate foamy viruses, including a new virus from a gorilla. *Virology* **207**, 217–228 (1995).
74. Schweizer, M. *et al.* Markers of foamy virus infections in monkeys, apes, and accidentally infected humans: appropriate testing fails to confirm suspected foamy virus prevalence in humans. *AIDS Res. Hum. Retroviruses* **11**, 161–170 (1995).
75. Goepfert, P. A. *et al.* Analysis of west African hunters for foamy virus infections. *AIDS Res. Hum. Retroviruses* **12**, 1725–1730 (1996).
76. Sandstrom, P. A. *et al.* Simian foamy virus infection among zoo keepers. *Lancet* **12**, 551–552 (2000).
77. McClure, H. M. Tumors in nonhuman primates: observations during a six-year period in the Yerkes primate center colony. *Am. J. Phys. Anthropol.* **38**, 425–429 (1973).
78. Seibold, H. R. & Wolf, R. H. Neoplasms and proliferative lesions in 1065 nonhuman primate necropsies. *Lab. Anim. Sci.* **23**, 533–539 (1973).
79. Willemssen, P. T. J. & de Graaf, F. K. Multivalent binding of K99 fimbriae to the N-glycolyl-GM<sub>3</sub> ganglioside receptor. *Infect. Immun.* **61**, 4518–4522 (1993).

 Online links

**DATABASES**  
The following terms in this article are linked online to:  
**LocusLink:** <http://www.ncbi.nlm.nih.gov/LocusLink>  
*BRCA1* | *BRCA2* | *CMAH* | *FOXP2* | *SIGLECL1* | *V10*  
**OMIM:** <http://www.ncbi.nlm.nih.gov/Omim>  
breast cancer | cystic fibrosis | multiple sclerosis | phenylketonuria | rheumatoid arthritis | systemic lupus | type II diabetes

**FURTHER INFORMATION**  
**Ajit Varki's laboratory:**  
<http://cmm.ucsd.edu/varki/biography.html>  
**Maynard Olson's laboratory:**  
<http://www.genome.washington.edu/UWGC>  
Access to this interactive links box is free online.