Comparative genomics of higher primates, including humans and Neandertals
(Svante Pääbo)

At a glance

The comparison of the human genome to the genomes of our closest living relatives, the apes, as well as our closest extinct relative, the Neandertal, will allow almost all genomic features specific to fully modern humans to be identified. Further studies of these genomes as well as functional tests in model systems will allow those features that have been functionally important during human evolution to be identified. Ultimately, this will uncover the biological basis for what made human cognition, language and culture possible and may provide inroads to a better understanding of diseases affecting human-specific traits.

Definition of research

The closest living evolutionary relatives of humans are the two chimpanzee species (“common chimpanzees” and bonobos) with whom we share a common ancestor about 5 to 7 million years ago. The other African ape, the gorilla, is almost as closely related to us, sharing a common ancestor 6 to 8 million years ago, while the common ancestor shared with the Asian orangutans lived some 12 to 16 million years ago.

Since our shared common ancestors with the chimpanzees, many extinct groups related to human ancestors, so-called hominins, existed. Of these, the ones most closely related to us were the Neandertals, with whom we share a common ancestor in the order of 400,000 to 500,000 years ago. They existed in western Eurasia until they became extinct about 30,000 years ago.

Genomic comparisons of humans with closely related apes and hominins are important not only in order to better understand humans origins and the biological background of what sets humans apart from other organisms, but also for understanding the basis for diseases or developmental problems that affect uniquely human traits, such as speech disorders, autism or schizophrenia.

Status of field

Technically, the field of genome sequencing is developing extremely rapidly. Very soon, it will be possible to sequence complete mammalian genomes to high accuracy at very little cost. This affects also other areas of genomics, for example high-throughput sequencing of RNA that now allows comparisons of levels of gene expression as well as patterns of spicing and transcription of parts of the genome other than the classical protein-coding genes. By comparison, mass spectroscopy-based high-throughput quantitative analysis of the proteome (i.e. all the proteins present in a tissue or a cell) and the metabolome (i.e. the small molecules produced by metabolic processes) is still in its infancy but in rapid development.

In general, these technical advances are first applied to humans, particularly medical questions and only slowly to other primates.

Currently, genome sequences of variable quality are available for humans, a chimpanzee, a gorilla, an orangutan and a rhesus macaque. The analysis of these genomes allows the identifications of millions of nucleotide changes that occurred on the human lineage since the divergence from the chimpanzee lineage as well as tens of thousands of insertions, deletions and duplications of pieces of DNA. The great challenge in the field is to identify those changes that form the biological basis for traits that are shared by all humans and distinguish them from all other species.
Important areas in this endeavor are comparisons of not only the DNA sequences of the genomes but also of functional aspects of the genomes, such as gene activity in the form of RNA and protein concentrations in different cells and organs and during various stages of development.

Recent technical advances have also made it possible to begin to sequence the genome of our closest extinct relative, the Neandertals, in order to identify those changes that occurred during the last few hundred of thousands of years when fully modern humans appeared.

**International activities**

Thanks mainly to efforts in the US, the human genome project and the sequences of a few ape genomes are currently being followed by determination of draft genome sequences of many primates. In the next few years, genome sequences of single representatives of almost all monkeys can be expected. MPG is contributing to this field by sequencing of the bonobo (pygmy chimpanzee) genome.

Several large projects are providing resources and insights into the genomic variation in humans (HapMap, ENCODE, 1000 Genomes). In the next few years these activities are expected to expand. No comparable efforts exist for the apes or other primates.

Efforts to sequence complete genomes of extinct animals are beginning to emerge thanks to adaptations of high-throughput DNA sequencing technologies. In this endeavor, the MPG has taken a lead with the Neandertal genome effort.

**Research opportunities and needs**

It will be important to more fully understand the genomic evolution of humans by both structural and functional comparisons of the human genome to the genomes of extant apes and extinct hominins. In order to achieve this, it will be necessary to obtain the same basic genomic resources that exist for humans in non-human primates, with a view to using those resources as the basis for subsequent evolutionary analyses, population genetics and disease gene identification.

To be useful, such comparisons will have to include not only comparisons of single genomes per species but also genomic variation within the species. This is important both because many groups are so closely related that they share variants between species and because many methods to identify genes that have been of importance for the species rest on comparisons of variation within species as well as differences between species.

The following areas will be particularly important in the next few years:

- Primate genomes of high and similar quality will need to be generated to allow differences between species to be defined accurately and comprehensively. Now, large and different numbers of errors in the different genomes make many analyses difficult. For example, currently the gorilla genome is of low quality and the chimpanzee X chromosome is sequenced to very low coverage. Over the next few years, it will be important to sequence all apes genomes to a quality similar to the human genome.
- Primate genomes need to annotated independently using cDNAs generated directly from the species in question. Currently, such annotations for all non-human primate genomes are simply “lifted over” from the human genome. This will allow differences in gene structure, splicing and regulatory RNAs to be detected.
- Primate genome variation panels that complement the "1000 genomes project" in humans will need to be created, at least for the great apes. This will allow the identification of genes that have been positively selected in humans as well as in the other primates. Currently the
largest studies performed non-human primates have sequenced only 100-200kb in approximately 20 individuals.

- To achieve the above goals it will be important to systematically collect blood samples and tissues from captive apes. Such material can be obtained when material is removed for medical reasons in primate sanctuaries.
- It will also be important to collect and establish genomic resources from wild apes. This is crucial since captive apes often represent only a limited sample of the total genetic variation in the species. The collection of such samples is all the more urgent as many of the wild populations can be expected to disappear over the next one or two decades. It is therefore necessary to systematically collect and preserve non-invasive samples from all great apes across their entire current habitats. This can be done in the form of collection of fecal and hair samples, and their preservation in ways that allow the retrieval of DNA in the future.
- From material collected from both captive and wild apes, immortalized cell lines and sequencing libraries that can be indefinitely replicated should be established whenever possible in order to provide genomic resources for the future. Methods should be developed to establish cell lines from novel sources of material, for example by establishing induced pluripotent stem (iPS) cells.

Neandertal genome sequencing will be continued with the aim of achieving a sequence of higher accuracy. It should be extended to other hominins whenever possible. In addition, the development and promotion of standards and protocols for working with ancient DNA, particularly ancient hominids, will be important to develop further.

- The analyses of the genomes of humans, Neandertals and apes will lead to the identification of features in the human genome that are candidates for being of importance for human-specific traits. Animal models for testing such candidate features are per definition not available since the traits of interest are unique to humans and most experiments cannot be done in apes (who cannot give informed consent) nor in Neandertals (who are extinct). Efforts to establish *in vitro* systems, for example in the form of tissue cultures that closely mimic physiological tissues from iPS cells, should therefore be pursued. Genetic manipulations of these *in vitro* systems and animal models in which genes or whole pathways are “humanized” will provide an avenue for testing aspects of traits specific to humans.

**Expected outcome and benefits**

The understanding of the biological basis for what made important aspects of human cognition, language and culture possible will be of substantial cultural interest. Since several common diseases affect aspects of human-specific traits, it is also expected that an understanding of the biological basis of these traits may provide an inroad into understanding their pathophysiology in addition to genetic association and linkage studies. This may apply to, for example, speech disorders, autism, and schizophrenia.

**References**


**Figures**

**Fig. 1**

![Fig. 1: Tree of humans and apes. (From: Kaessmann, H., and Pääbo, S.: The genetical history of humans and the great apes. Journal of Internal Medicine 251: 1-18 (2002)).](image-url)
Fig. 2: A skeleton of a Neandertal and a human. (Photo: K. Mowbray, Reconstruction: G. Sawyer and B. Maley, Copyright: Ian Tattersall)

Fig. 3: Artistic view of murine model systems.