https://doi.org/10.1038/d41586-025-00912-8

News & views

Comparative genomics

Evolutionary insights from complete ape genomes

Lukas Kuderna

Scientists have fully sequenced the genomes of six living ape species, enabling long-awaited comparisons of hard-to-assemble genomic regions.

Apes, the group of primates that includes humans, are our closest evolutionary relatives. Comparisons between the genomes of humans and those of other apes have been crucial for understanding the function of the human genome and our own evolutionary history. But, because ape genomes are large and contain repetitive sequences, many genomic regions have been difficult to sequence and reconstruct accurately, which has so far resulted in incomplete representations that preclude full comparisons. Writing in *Nature*, Yoo *et al.*[1] report essentially complete genome sequences for six ape species that represent all of the main ape lineages: chimpanzee (*Pan troglodytes*), bonobo (*Pan paniscus*), gorilla (*Gorilla gorilla*), Bornean orangutan (*Pongo pygmaeus*), Sumatran orangutan (*Pongo abelii*) and siamang (*Symphalangus syndactylus*).

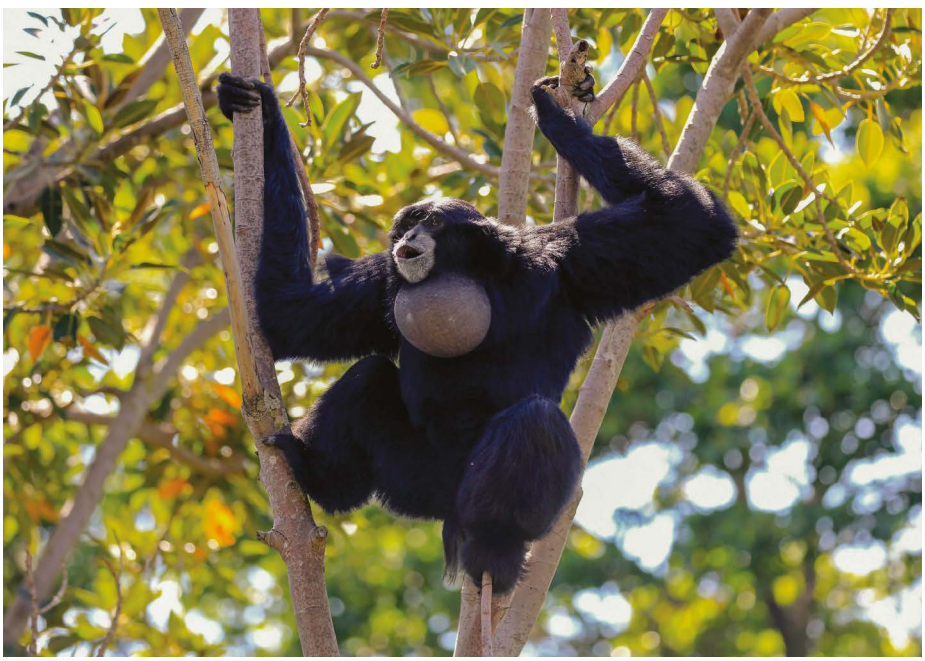
Shortly after the first human genome sequence was finalized in 2003, a chimpanzee assembly was released[2]. This was followed by assemblies for other great apes, such as the gorilla[3], Sumatran orangutan[4] and bonobo[5], and small apes that are less closely related to humans than are great apes[6]. These genomes offered a valuable opportunity to catalogue the genetic differences that have accumulated during the evolution of apes, including changes that are unique to humans. But, because these initial releases were incomplete drafts, comparisons could be made only between properly resolved portions of the genome. These studies therefore focused only on relatively small differences, and excluded extremely repetitive sequences and large-scale structural differences, such as inversions and duplications of genomic sequences.

Yoo and colleagues' work elevates the quality of ape genome sequences to the same level as existing sequences for humans, enabling the authors to investigate the evolutionary history of almost the entire human genome. This invaluable resource also aids comparative analyses of previously inaccessible genomic regions, many of which have biomedical relevance.

The newly characterized regions include the parts of chromosomes that are responsible for guiding cell division. These regions, known as centromeres, separate each chromosome into two arms. Centromeres are composed of small repetitive sequences called α-satellites, and patterns of α-satellite repeats can themselves become repeated in what are known as higher-order arrays. Although individual centromeres can amount to millions of nucleotide bases, the sequence organization of higher-order arrays was not well understood.

In their study, Yoo *et al.* characterize the complete centromere composition of most chromosomes across apes. They describe extensive variations in length and sequence composition between and within species, which are partly the result of fast and recent evolution. For example, around 40% of the centromeres in bonobos decreased in size by about 300 times, which happened at most one million years ago, resulting in 'mini-centromeres' that are specific to that lineage. Similarly, despite the relatively recent divergence between Bornean and Sumatran orangutans (about 960,000 years ago), around one-fifth of the chromosomes of Bornean orangutans contain newly emerged higher-order arrays, whereas those of Sumatran orangutans do not, exemplifying the changes in these regions over short evolutionary periods.

Some chromosomes (13, 14, 15, 21 and 22 in humans) are acrocentric, which means the centromere is close to one end of the chromosome. The short arms of acrocentric chromosomes in humans carry almost no genes – except for those that encode ribosomal RNA, which is needed to build ribosomes, the molecular machines that synthesize proteins. Numerous copies of ribosomal RNA genes are found in large arrays in nucleolar organizer regions (NORs), which are surrounded by repetitive sequences. Because of their importance, ribosomal RNA genes are



JURGEN & CHRISTINE SOHNS/GETTY

Figure 1 | A siamang (*Symphalangus syndactylus*).

evolutionarily conserved, but Yoo *et al.* show that they are highly variable in copy number and describe widespread structural variation in the surrounding regions, often resulting in a complete change of sequence.

The authors also observe that the number of NORs differs across species. For example, the siamang (Fig. 1) has only one, whereas there are ten in the orangutans. NORs have changed location throughout ape evolution, meaning that the chromosomes they are found on differ across species. This could be because of a process called heterologous recombination (an unequal exchange of DNA between two chromosomes or regions), which can help to maintain the sequence and function of NORs by removing the differences between copies of duplicated regions. The result of this is a phenomenon known as concerted evolution, whereby different copies of a gene evolve in similar ways. The authors show that, within the same species, concerted evolution probably happens between different NORs, and also between copies of ribosomal RNA genes in the same array, which are more similar to each other than to the equivalent genes in other NORs. These observations exemplify the value of the resources that the authors have generated, providing a toolkit to describe and understand the evolution of many vastly diverse and divergent genomic regions.

Yoo and colleagues also analysed segmental duplications. These gene-rich duplicated regions of the genome can be thousands to millions of bases in length and contain copies of genes that have high sequence similarity. Segmental duplications are important because they underlie several human diseases and have been key to shaping the evolution of great apes. For example, human-specific copy-number increases of some genes are thought to be involved in regulating the expansion in the volume of our frontal cortex, the brain region responsible for higher cognitive functions. A direct sequence-level comparison enabled Yoo and colleagues to identify the structure and genes in duplicated regions, and to revisit the timing and rate of the expansions of these regions across different ape ancestors.

Duplications can change gene expression by regulating the amount of gene product that is made, but they also produce redundant copies of a gene. Redundant gene copies were theorized[7] to be the substrate for evolutionary innovations, being able to give rise to proteins with differing functions as they acquired alterations in their sequences. An alternative theory[8] proposes that changes to gene regulation, rather than sequence, underlie the differences between species. For example, humans and our closest living relatives, chimpanzees, have protein-coding genes that are remarkably similar, so, according to this theory, it seems unlikely that variations in sequence could account for the substantial organismal differences.

Interestingly, the authors found hundreds of genes with copy-number expansions specific to certain ape lineages, leading them to conclude that these could be a potentially underappreciated source of functional innovation, which might challenge the idea that it was mainly changes to gene regulation that drove speciation (the formation of distinct species) in apes. Conversely, they find that the ancestral sequences that have diverged most quickly in humans are enriched in different types of gene-regulatory element — non-protein-coding sequences that regulate gene expression. This apparent contradiction could support either theory, but it also highlights some of the opportunities left for further exploration. Although highly accurate genome sequences are, without a doubt, a crucial resource for studying the evolution of challenging genomic regions in apes, they are just the first step towards studying the possible functional consequences of those changes. Massive efforts have been made to catalogue gene-regulatory elements in humans[9,10], but comparable resources for other apes are scarce at best.

There are other questions remaining. Many of the newly characterized regions show extreme sequence variation not only between species but also within them. The latter observation is mostly based on a comparison of maternally and paternally inherited DNA from one individual, so it might offer just a glimpse of the true extent of variation. Furthermore, about 0.1–0.8% of bases across all six species could not be accurately assembled, and about 20% of chromosomes do not yet have fully uninterrupted sequences. An acceptable threshold for what can be considered a 'complete' genome is a moving target chased by scientific innovation — the human genome was brought to comparable quality only three years ago[11].

Nevertheless, Yoo and colleagues bring great improvements in sequence resolution and accuracy. The characterization of the structure and variability of newly resolved sequences of these effectively complete genomes will pave the way to a deeper and more refined understanding of human evolution in the context of apes, not by delivering definitive answers, but by laying the foundation for countless further research avenues.

Lukas Kuderna is in the Illumina Artificial Intelligence Laboratory, Illumina, San Diego, California 92122, USA.
e-mail: lkuderna@illumina.com

1. Yoo, D. *et al. Nature* https://doi.org/10.1038/s41586-025-08816-3 (2025).
2. The Chimpanzee Sequencing and Analysis Consortium. *Nature* **437**, 69–87 (2005).
3. Scally, A. *et al. Nature* **483**, 169–175 (2012).
4. Locke, D. P. *et al. Nature* **469**, 529–533 (2011).
5. Mao, Y. *et al. Nature* **594**, 77–81 (2021).
6. Carbone, L. *et al. Nature* **513**, 195–201 (2014).
7. Ohno, S. *Evolution by Gene Duplication* (Springer, 2014).
8. King, M.-C. & Wilson, A. C. *Science* **188**, 107–116 (1975).
9. ENCODE Project Consortium. *Nature* **489**, 57–74 (2012).
10. GTEx Consortium. *Nature* **550**, 204–213 (2017).
11. Nurk, S. *et al. Science* **376**, 44–53 (2022).

The author declares competing interests; see go.nature.com/3rnmk2r for details.