# Chromosome components central to cell division evolve rapidly

Regions of chromosomal DNA called centromeres are crucial to dividing cells. Centromere sequences from a human genome have been fully characterized and compared with those from other humans and non-human primates, revealing dynamic and rapid patterns of mutational change that will improve understanding of centromere evolution and function.

**This is a summary of:**
Logsdon, G. A. *et al*. The variation and evolution of complete human centromeres. *Nature* **629**, 136–145 (2024).

**Cite this as:**
*Nature* https://doi.org/10.1038/d41586-024-02523-1 (2024).

## The problem

Centromeres are chromosomal regions that are crucial for ensuring accurate segregation of copies of genetic material during cell division. However, centromeres have been challenging to sequence and assemble completely, because they are made up of nearly identical sequences that are large and extremely repeat-rich compared with other parts of the human genome. Thus, they have typically been excluded from sequencing studies, limiting researchers' understanding of genetic variation and how it affects centromere function. Although centromeres have long been known to evolve rapidly, a lack of sequencing data from centromeres from multiple humans and closely related ape species has resulted in an incomplete understanding of their evolution.

## The solution

Our approach takes advantage of the complementary strengths of two technologies that can resolve genomic sequences by reading relatively long stretches of DNA. The first, produced by the California-based company Pacific Biosciences, uses high-fidelity sequencing to provide the high base accuracy needed to produce the backbone of the assembly. The second — ultra-long sequencing, developed by the UK firm Oxford Nanopore Technologies — provides the very long sequence reads that are necessary to reach across centromeres, which are often millions of base pairs long. Thus, we resolved the DNA sequence and epigenetic features (modifications of the DNA–protein complex chromatin that do not alter DNA sequence) of all centromeres in the chromosomes of a human cell line called CHM1, which has a duplicated paternal genome and no maternal genome. We also obtained the full DNA sequence of centromeres in a subset of six chromosomes (5, 10, 12, 20, 21 and X) from four primate species, including humans.

When comparing CHM1 centromeres with their counterparts in the first completely sequenced human genome (from a similar cell line called CHM13)[1–3], we found that over one-fifth of them vary in length by more than 1.5-fold, and that one-third show distinctly different organization in regions called α-satellite higher-order repeat (HOR) arrays. Furthermore, we found that the position of the kinetochore — the protein structure that connects centromeric chromatin and the spindle structure on which chromosomes align during cell division — differs in location by more than 500,000 base pairs in about 26% of centromeres. We compared the sequenced centromeres with a subset of those from humans and other primate species — chimpanzee (*Pan troglodytes*), orangutan (*Pongo abelii*) and macaque (*Macaca mulatta*) — and identified characteristics that are specific to each primate lineage. We estimate that, compared with non-repetitive regions of the human genome, centromeres mutate at least 4.1 times faster at their peripheries and probably much faster in their interiors. We identify the emergence of evolutionarily novel centromeric sequences and structures in subsets of human chromosomes that, in some cases, associate with chromatin at the kinetochore-binding position (Fig. 1).

## Future directions

The ability to completely sequence centromeres will enable researchers to estimate the mutation rate of centromeres, and provide insight into the mutational processes that drive such dynamic structural changes. The speed and magnitude of this change raises the question of how these regions can so stably confer function despite constant genomic turnover. This centromere paradox has long remained unsolved[4], but our study begins to shed light on the range of sequence structures that are compatible with centromere function.

This work provides a first glimpse of the variation between two complete sets of human centromeres, but there is more to be discovered. For example, we still do not know the full extent of centromere variation in healthy individuals and whether this is different in disease contexts. Furthermore, it is unclear whether centromeres vary in different tissues, throughout development and ageing and across generations.

In the future, we hope to develop a baseline of human centromere variation and a model for its evolution in primates. Efforts are under way to sequence and assemble near-complete genomes, including centromeres, from a diverse sample of hundreds of people[5]. Using these assemblies, we plan to build a road map of healthy centromere structure and function that can be used to identify aberrant centromeres that are associated with disease and disease susceptibility.

**Glennis A. Logsdon** is at the University of Pennsylvania, Philadelphia, Pennsylvania, USA, and **Evan E. Eichler** is at the University of Washington School of Medicine, Seattle, Washington, USA.

## EXPERT OPINION

"Centromeres are essential for chromosome segregation and are frequently composed of complex tandem repeat arrays. Until recently, these centromeric arrays could not be reliably assembled. In this work, the authors present centromere-complete assemblies and use a comparative genomics approach to infer new insights into the genetic and epigenetic organization of human and primate centromeres, with implications for how they evolve." (CC BY 4.0)

**Ian Henderson** is at the University of Cambridge, Cambridge, UK.

## FIGURE

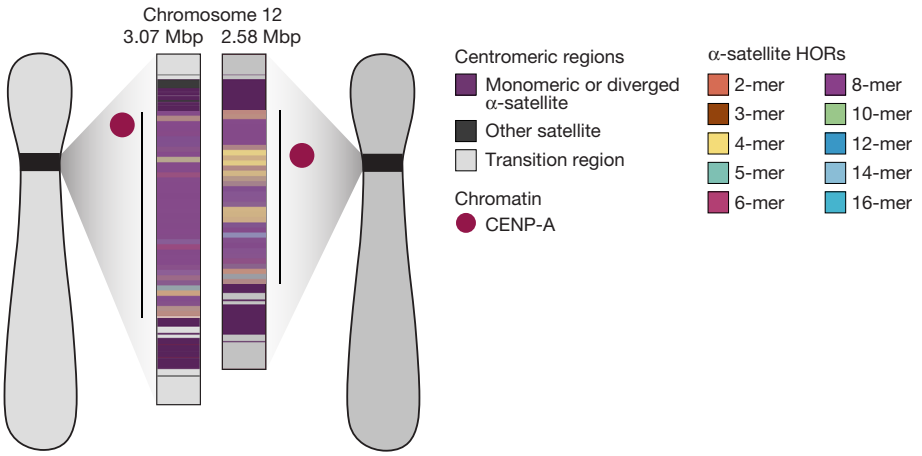


**Figure 1 | Variation of the centromere in chromosome 12 from two humans.** Chromosomal sequences called centromeres contain long, non-coding arrays of repeat units known as α-satellite higher-order repeats (HORs). A comparison of these arrays in the chromosome 12 centromeres of the human cell lines CHM1 (left) and CHM13 (right) shows that the CHM13 array has an abundance of α-satellite HORs made up of 4-mers (four α-satellite monomer HORs; yellow), whereas the CHM1 array does not. Typically, α-satellite HOR arrays are the site of the kinetochore – the proteinaceous structure responsible for aligning chromosomes during cell division. Dark red circles show the location of CENP-A, a component of chromatin (a complex of DNA and proteins) that determines the position of the kinetochore. Mbp, length in megabase pairs. Logsdon, G. A. *et al.*/*Nature* (CC BY 4.0).

## BEHIND THE PAPER

Throughout this project, we examined hundreds of human centromeres and compared their genetic and epigenetic landscapes with each other as well as with those from closely related species. What we discovered is that no two sets of centromeres — even those from closely related individuals — are identical, and that centromeric regions can, in essence, serve as a 'genomic fingerprint' of each human. Each fingerprint has its own evolutionary trajectory, whereby dynamic changes in structure and content rapidly diversify the sequence through a mutational mechanism that is, as yet, poorly understood. This realization has opened our eyes to the importance of understanding the epigenetic, genetic and evolutionary variation of this last frontier of the human genome.

**G.A.L.**

## REFERENCES

1. Nurk, S. *et al. Science* **376**, 44–53 (2022).
2. Altemose, N. *et al. Science* **376**, eabl4178 (2022).
3. Logsdon, G. A. *et al. Nature* **593**, 101–107 (2021).
4. Henikoff, S., Ahmad, K. & Malik, H. S. *Science* **293**, 1098–1102 (2001).
5. Liao, W.-W. *et al. Nature* **617**, 312–324 (2023).

## FROM THE EDITOR

Until recently, the genomic features of centromeres were difficult to study because of the complex and repetitive nature of the centromere DNA sequence. In this paper, assembling all the centromeres from a human cell line has allowed the authors to perform a first analysis of genetic variation across human centromeres. The comparative analysis with complete centromeres from three non-human primate species also provides important insights into the evolution of centromeres in the primate lineage.

**Michelle Trenkmann**, Senior Editor, *Nature*