

# EXPLAINED

Relevant for: null | Topic: Biology related Matters

To enjoy additional benefits

CONNECT WITH US

May 21, 2023 10:43 pm | Updated May 22, 2023 08:21 am IST

COMMENTS

SHARE

READ LATER

For representative purposes. | Photo Credit: Getty Images

**The story so far:** A new study published in the May 10 issue of the *Nature* journal describes a pangenome reference map, built using genomes from 47 anonymous individuals (19 men and 28 women), mainly from Africa but also from the Caribbean, Americas, East Asia, and Europe.

The genome is the blueprint of life, a collection of all the genes and the regions between the genes contained in our 23 pairs of chromosomes. Each chromosome is a contiguous stretch of DNA string. In other words, our genome consists of 23 different strings, each composed of millions of individual building blocks called nucleotides or bases. The four types of building blocks (A, T, G and C) are arranged and repeated millions of times in different combinations to make all of our 23 chromosomes. Genome sequencing is the method used to determine the precise order of the four letters and how they are arranged in chromosomes. Sequencing individual genomes helps us understand human diversity at the genetic level and how prone we are to certain diseases.

The genome is an identity card like Aadhaar. As each of our Aadhaar card is unique, so is our genome. As sequencing individual genomes of all humans is expensive, we do not yet have all our genome identity cards. To circumvent this, one can have a collective identity card. For example, we can have a single genome identity card for everyone living in a region.

When genomes are newly sequenced, they are compared to a reference map called a reference genome. This helps us to understand the regions of differences between the newly sequenced genome and the reference genome. One of this century's scientific breakthroughs was the making of the first reference genome in 2001. It helped scientists discover thousands of genes linked to various diseases; better understand diseases like cancer at the genetic level; and design novel diagnostic tests. Although a remarkable feat, the reference genome of 2001 was 92% complete and contained many gaps and errors. Additionally, it was not representative of all human beings as it was built using mostly the genome of a single individual of mixed African and European ancestry. Since then, the reference genome map has been refined and improved to have complete end-to-end sequences of all the 23 human chromosomes.

Although complete and error-free, the finished reference genome map does not represent all of human diversity. The new study published in *Nature* changes this. The main paper and the accompanying articles published in the same journal and *Nature Biotechnology* describe the making of the pangenome map, the genetic diversity among the 47 individuals, and the

computational methods developed to build the map and represent differences in those genomes.

Unlike the earlier reference genome, which is a linear sequence, the pangenome is a graph. The graph of each chromosome is like a bamboo stem with nodes where a stretch of sequences of all 47 individuals converge (similar), and with internodes of varying lengths representing genetic variations among those individuals from different ancestries. To create complete and contiguous chromosome maps in the pangenome project, the researchers used long-read DNA sequencing technologies, which produce strings of contiguous DNA strands of tens of thousands of nucleotides long. Using longer reads helps assemble the sequences with minimum errors and read through the repetitive regions of the chromosomes which are hard to sequence with short-read technologies used earlier.

Although any two humans are more than 99% similar in their DNA, there is still about a 0.4% difference between any two individuals. This may be a small percentage, but considering that the human genome consists of 3.2 billion individual nucleotides, the difference between any two individuals is a whopping 12.8 million nucleotides. A complete and error-free human pangenome map will help us understand those differences and explain human diversity better. It will also help us understand genetic variants in some populations, which result in underlying health conditions. The pangenome reference map has added nearly 119 million new letters to the existing genome map and has already aided the discovery of 150 new genes linked to autism.

Although the project is a leap forward, genomes from many populations are still not a part of it. For example, genomes from more people from Africa, the Indian sub-continent, indigenous groups in Asia and Oceania, and West Asian regions are not represented in the current version of the pangenome map.

Even though the current map does not contain genome sequences from Indians, it will help map Indian genomes better against the error-free and complete reference genomes known so far. Future pangenome maps that include high quality genomes from Indians, including from many endogamous and isolated populations within the country, will shed light on disease prevalence, help discover new genes for rare diseases, design better diagnostic methods, and help discover novel drugs against those diseases.

*Binay Panda is a Professor at the Jawaharlal Nehru University, New Delhi*

COMMENTS

SHARE

[Text and Context](#) / [The Hindu Explains](#) / [genetics](#) / [science and technology](#) / [journals and diaries](#)

BACK TO TOP

Comments have to be in English, and in full sentences. They cannot be abusive or personal. Please abide by our [community guidelines](#) for posting your comments.

We have migrated to a new commenting platform. If you are already a registered user of The Hindu and logged in, you may continue to engage with our articles. If you do not have an account please register and login to post comments. Users can access their older comments by logging into their accounts on Vuukle.

END

CrackIAS.com