

ABSTRACTSCIENCE

Building the most exhaustive family tree

Shubashree Desikan

MARCH 02, 2022 10:30 IST

UPDATED: MARCH 02, 2022 19:22 IST

A recent study constructs a genealogy of all humans from 27 million fragments of ancestral genomes



Andrew Wilder Wohns et al, “A unified genealogy of modern and ancient genomes”, *Science*, Vol. 375, 2022. DOI: [10.1126/science.abi8264](https://doi.org/10.1126/science.abi8264)

Building family trees is very popular with many people. But such an attempt is likely to take the process back to just a few generations at the most. In some cases, families have been traced back to hundreds of years. Now, we are to see that science has enabled us to take this process back thousands of years, and we can have a family tree of the entire human race, according to a study published in the *Science* journal. These family trees are constructed not by identifying individuals that one has descended from, but by taking from the human genomic databases, details of each chromosome (autosome) and constructing a “tree” that relates it to the parent from whom that particular chromosome was inherited. New work by researchers from the University of Oxford’s Big Data Institute builds up the family tree of the entire human race using this new methodology. This is the largest family tree ever, built with about 27 million ancestral haplotype fragments (ancestral DNA). The research has been published in the February 25 issue of *Science*, this

year (“A unified genealogy of modern and ancient genomes”, Andrew Wilder Wohns et al, *Science*, Vol. 375, 2022).

Inferring historical data

Among existing methods to build up a genealogy of humans, or any other organism whose genome data exists for that matter, the one involving “trees” appears to be more accurate. We already have such trees that have combined thousands of genomes and traced out demographic data. According to this study, ancient genomes can be integrated into these trees to give data from which one can infer historical details of demography. The present study, the latest in this line, takes into account more than 3,500 high-quality and ancient genomes from more than 215 different human populations. In addition, the researchers used, more than 3,000 ancient genomes to improve inferences from the trees. A “Perspective” article in the same issue of *Science* (written by Jasmin Rees and Aida Andres) gives an easy-to-read gist of the paper.

Combining datasets that have been built up independently poses a huge challenge. The authors refer to discrepancies between cohorts due to errors, differing sequencing techniques, and the processing of variants, which can lead to a lot of “noise” that can drown out the “signal.” However, they do a smart building up of this tree, the largest one built up so far, and reproduce several well-known demographic results that have been given by previous methods such as using mitochondrial DNA and Y chromosome analyses. They reiterate the out-of-Africa emergence result, for one thing.

The paper proposes a new method to determine the ancient geographical location of evolutionary events. They take the mid-points of the geographical locations of individuals featured in the tree and postulate that the location of the parents is at the mid-point of these. Thereafter, by repeating this procedure, they arrive at the location of the theoretical earliest common ancestor. Wohns et al write that despite the fact that the geographical centre of the sampled individuals is in Central Asia, when you go back 72,000 years into the past, this geographical centre shifts to Northeast Africa “and remains there until the oldest common ancestors are reached.” They further state, “the geographic centre of the 100 oldest ancestral haplotypes (which have an average age of approximately 2 million years) is located in Sudan.” While these results are supported by the oldest fossil findings, the authors caution that if the data had consisted of a grid sampling of genomes from Africa, for instance, the centre of gravity, which is now Sudan, could shift.

Limitations of the study

The “Perspective” article also outlines some of the limitations of the tree method, namely, that there is an uncertainty in evolutionary parameters and errors in reading ancient genomes, but also point out that with more and better-quality data coming in, the tree

method is likely to provide good answers; as larger and more data from under-represented populations becomes available, the answers would become more accurate.

While this particular study has focussed on human genealogy, the authors point out that this can be used to trace out the ancestry of any species for which genomic data exists.

It is important to appreciate that such an analysis would not have been possible in the twentieth century, when the human genome had not been sequenced, cutting edge techniques for isolating, identifying and sequencing genomes of archaic sources had not been developed, sophisticated computer simulation methods were not available and such perspectives had not been established. So truly, the building of this huge genealogy is a feat of twenty-first century science, and it brings together evolutionary biologists, repositories of information, simulation experts and high-power computing and data science to yield hitherto unimagined results.

THE GIST

A new study by researchers from Oxford has enabled us to see a family tree of the entire human race. These family trees are constructed by taking from the human genomic databases, details of each chromosome (autosome) and constructing a “tree” that relates it to the parent from whom that particular chromosome was inherited.

According to this study, ancient genomes can be integrated into these trees to give data from which one can infer historical details of demography. It takes into account more than 3,500 high-quality and ancient genomes from more than 215 different human populations.

Without cutting edge techniques for isolating, identifying and sequencing genomes of archaic sources and sophisticated computer simulation methods of 21st century science, such a feat would have been impossible.

THE HINDU

Our code of editorial values

1. Comments will be moderated by The Hindu editorial team.
2. Comments that are abusive, personal, incendiary or irrelevant cannot be published.
3. Please write complete sentences. Do not type comments in all capital letters, or in all lower case letters, or using abbreviated text. (example: u cannot substitute for you, d is not 'the', n is not 'and').
4. We may remove hyperlinks within comments.
5. Please use a genuine email ID and provide your name, to avoid rejection.

Printable version | Mar 2, 2022 7:38:31 pm | <https://www.thehindu.com/sci-tech/science/building-the-most-exhaustive-family-tree/article65143893.ece>

© THG PUBLISHING PVT LTD.

