Inferring human evolutionary history

Unified genetic genealogy improves our understanding of human evolutionary history

By **Jasmin Rees** and **Aida Andrés**

Genomes are invaluable tools to infer the demographic and adaptive history of human populations, including migrations, population splits, admixture, and genetic adaptations. Growing datasets of modern and ancient genomes make this possible, but their massive size comes with important challenges, demanding techniques that analyse massive amounts of data in reasonable amounts of time while using as much information as possible. Combining genomes from different datasets poses perhaps an even greater challenge, especially when it comes to integrating ancient and modern genomes. On page XXX of this issue, Wohns *et al.* (*1*) surmount some of these challenges to construct the largest human genealogy to date, integrating modern and ancient genomes from multiple datasets to infer key events in human history together with their timing and geographical location.

Phylogenetic trees are used to represent the evolutionary or genetic relationships among species. Similar trees can represent the relationships among individuals within a species, analogous to how family genealogies represent relationships between family members. However, due to the effects of recombination, each locus has a slightly different evolutionary history and therefore a different tree. Tree-recording methods (*2*) can infer these trees along the genome, with each tree representing what can be considered a nearly complete history of the locus. As a consequence, tree-recording is superior to most classical methods, which condense complicated evolutionary patterns in relatively simple summary statistics. By combining trees across the genome, a theoretical genealogy can be generated that embodies the genetic relationships among sampled individuals and their inferred ancestors, within and across populations. If the sampling of genomes was sufficiently comprehensive, such genealogy would in theory represent relationships over the entire species, capturing the genetic history of modern humans today.

Owing to a series of impressive theoretical and computational advances, genetic trees can now be inferred for thousands of genomes (*3-7*). Analyses of these trees have already improved our knowledge of human demography (*4-5, 8-9*) and promise to improve our ability to identify targets of natural selection (*5, 8, 10-12*). Further, high-quality ancient genomes can be directly integrated in the tree, greatly increasing certainty of evolutionary processes by supporting inferences with data from the past. This is particularly helpful to study populations that are extinct or contributed little to modern populations, such as Neanderthals and Denisovans (archaic humans; *13-14*), and tree-based studies are helping us better understand the nature of interbreeding between modern and archaic groups (*5, 8-9*).

Unfortunately, the DNA in ancient remains is usually highly degraded and the vast majority of ancient genomes are not of sufficient quality to be fully incorporated into the trees. Wohns *et al.* present an approach that aims to skirt inaccuracies introduced by ancient DNA by using such ancient genomes only to help time the emergence of alleles. This allows the use of hundreds of ancient genomes whilst limiting the effects of errors.

Wohns *et al.* thus generate an impressive, unified human genealogy from >3500 modern and high-quality ancient genomes from over 215 different human populations, using over 3000 additional ancient genomes to improve inferences from the trees. With this unified genealogy, key events in history such as population size changes, splits or migrations become strikingly apparent. They identify well-resolved events such as the Out-of-Africa migration and suggest multiple severe reductions in population size through human history.

Further, Wohns *et al.* propose a novel method that, applied to the unified genealogy, allows them to estimate ancestral geographical location of evolutionary events. While efforts to place past events geographically are not unique (*15*), Wohns *et al.* use a non-parametric method to estimate the theoretical location of inferred human ancestors simply as the midpoint of the geographic coordinates of descendants – with these calculations extended backwards up the tree to reach a theoretical common ancestor of all individuals. Using this method, Wohns *et al.* infer an average ancestral location of all sampled humans in Northeast Africa by 72kya and until the oldest common ancestors of all individuals. This simple method works well to refine known ancestral locations and, as sampling improves, it has the potential to identify currently unknown human movements. More generally, the trees generated in this study will prove undoubtedly useful to those studying human evolution.

Whilst tree-recording methods represent, as demonstrated by Wohns *et al.*, an exciting and promising avenue of harmonising datasets across time and space, they are not without limitations. There remains uncertainty in evolutionary parameters and ancient genomes, and it is unlikely that the use of low-quality ancient DNA will ever be without a degree of error, except when using rare high-quality genomes or if inferring missing data becomes remarkably accurate. Still, as genomic datasets continue to grow, genealogical methods will be increasingly useful to represent the wealth of genomic data. Perhaps most importantly, as larger numbers of modern and ancient genomes from under-represented populations become available, our understanding of human demography, currently still biased to well-sampled populations, will improve both in detail and in scope.

The power and resolution of tree-recording methods promise to help clarify the evolutionary history of humans and other species, and it is likely that going forward the most powerful ways to infer evolutionary history will have their foundations firmly set in these methods.

1UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, London, UK. 2Genetics and Genomic Medicine Programme, Great Ormond Street Institute of Child Health, University Collge London, London, UK

REFERENCES AND NOTES
1. Wohns *et al.*, *Science* **375**, XXX (2022).
2. M. Rasmussen *et al.*, *PLoS Genet.* **10**, 5 (2014).
3. J. Kelleher *et al.*, *PLoS Comput.l Biol.* **12**, 1 (2016).
4. J. Kelleher *et al.*, *Nat. Gen.* **51**, 1330 (2019).
5. L. Speidel *et al.*, *Nat. Gen.* **51**, 1321, (2019).
6. K. Harris *et al.*, *Nat. Gen.* **51**, 1306 (2019).
7. P. Ralph *et al.*, *Genetics* **215**, 779 (2020).
8. L. Speidel *et al.*, *Mol. Biol. Evol.* **38**, 3497 (2021).
9. N. Schaefer *et al.*, *Sci. Adv.* **7**, 776 (2021).
10. Y. Field *et al.*, *Science* **354**, 760 (2016).
11. A. Stern *et al.*, *PLoS Genetics* **15**, 9 (2019).
12. J. Stern *et al.*, *Am. J. Hum. Genet.* **108**, 219 (2021).
13. .M. Meyer *et al.*, *Science* **338**, 222 (2012).
14. .K. Prüfer *et al.*, *Nature* **505**, 43 (2014).

15. M. Osmond *et al.*, *bioRxiv*, https://doi.org/10.1101/2021.07.13.452277 (2021).

Figure Title: A simplification of the genealogy reconstruction and geographical inference from Wohns *et al.*

Figure Legend: Schematic representation of the metholodology, from left to right: 1) building the tree from modern human samples (black) and high-quality ancient genomes (orange), showing the inferred ancestral humans (grey), 2) using additional ancient human samples (red) to help infer the ages of alleles, 3) infering the geographical location of inferred ancestors (dashed lines), using ancient and modern samples of known location (solid lines).