

A weakly structured stem for human origins in Africa*

Ragsdale, A.P., Weaver, T.D., Atkinson, E.G. et al. A weakly structured stem for human origins in Africa. Nature 617, 755–763 (2023)

**Ancient DNA from Eurasia has helped to clarify early human history outside Africa, but there is no comparably ancient DNA to elucidate early history in Africa.*

Motivation

Model misspecification

Bottlenecks in progress

Weaknesses in previous studies

Demographics

- *Nama* (Khoe-San from South Africa)
- *Mende* (from Sierra Leone; from phase 3 of the 1000 Genomes Project)
- *Gumuz* (recent descendants of a hunter-gatherer group from Ethiopia)
- *Amhara* and *Oromo* (eastern African agriculturalists from Ethiopia)
- *British* individuals (1000 Genomes Project)
→ representative source of back-to-Africa gene flow and recent colonial admixture in South Africa

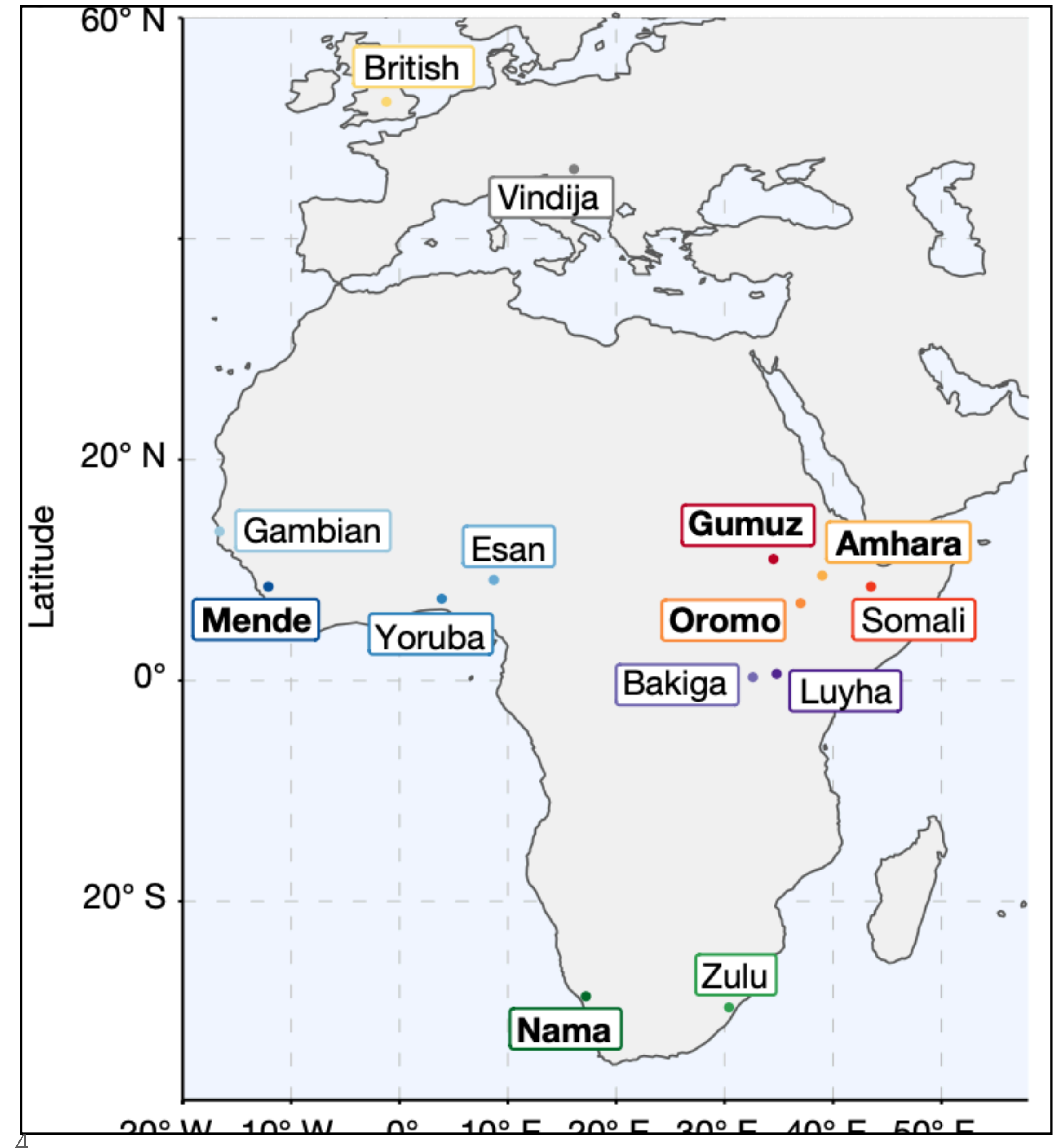


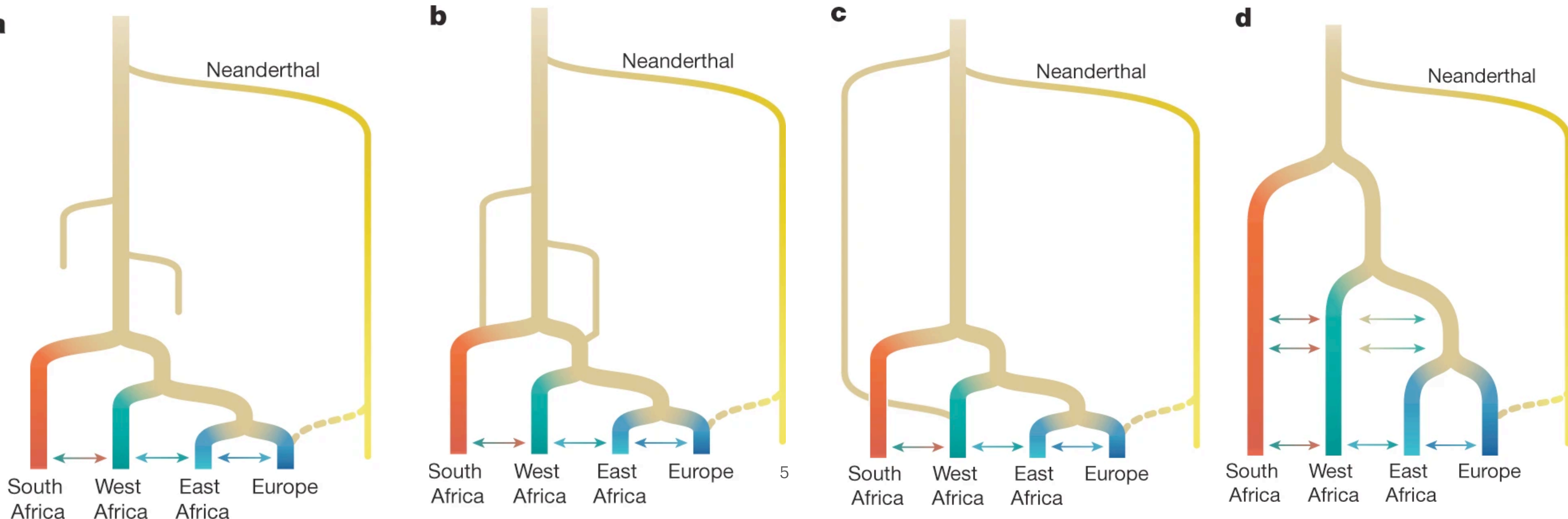
Fig 2a, Ragsdale e.al. (2023)

Starting Models

Recent expansion

Recent expansion
with regional persistence

Archaic admixture African multiregional



A maximum-likelihood inference framework helped gauge the relative time depth of divergence, assuming that Neanderthals diverged 550,000 years ago from a common stem.

Methods

Data and Sequencing

- African Diversity Reference Panel (ADRP) and 1000 Genomes Project (1KGP) populations.
- Merged with the high-coverage *Neanderthal* genome from Vindija Cave.
- Variant removal in high linkage disequilibrium: **ADMIXTURE** and **PCA** ($R^2 > 0.1$).

Methods

Linkage disequilibrium and diversity statistics

- Expectation & likelihood-based parameter optimisation → software: **Moments**
- **Testing multiple scenarios** of early population structure:
 1. long-lasting continuous migration between stem populations
 2. periods of isolation with subsequent merger events

Methods

Model specification and fitting

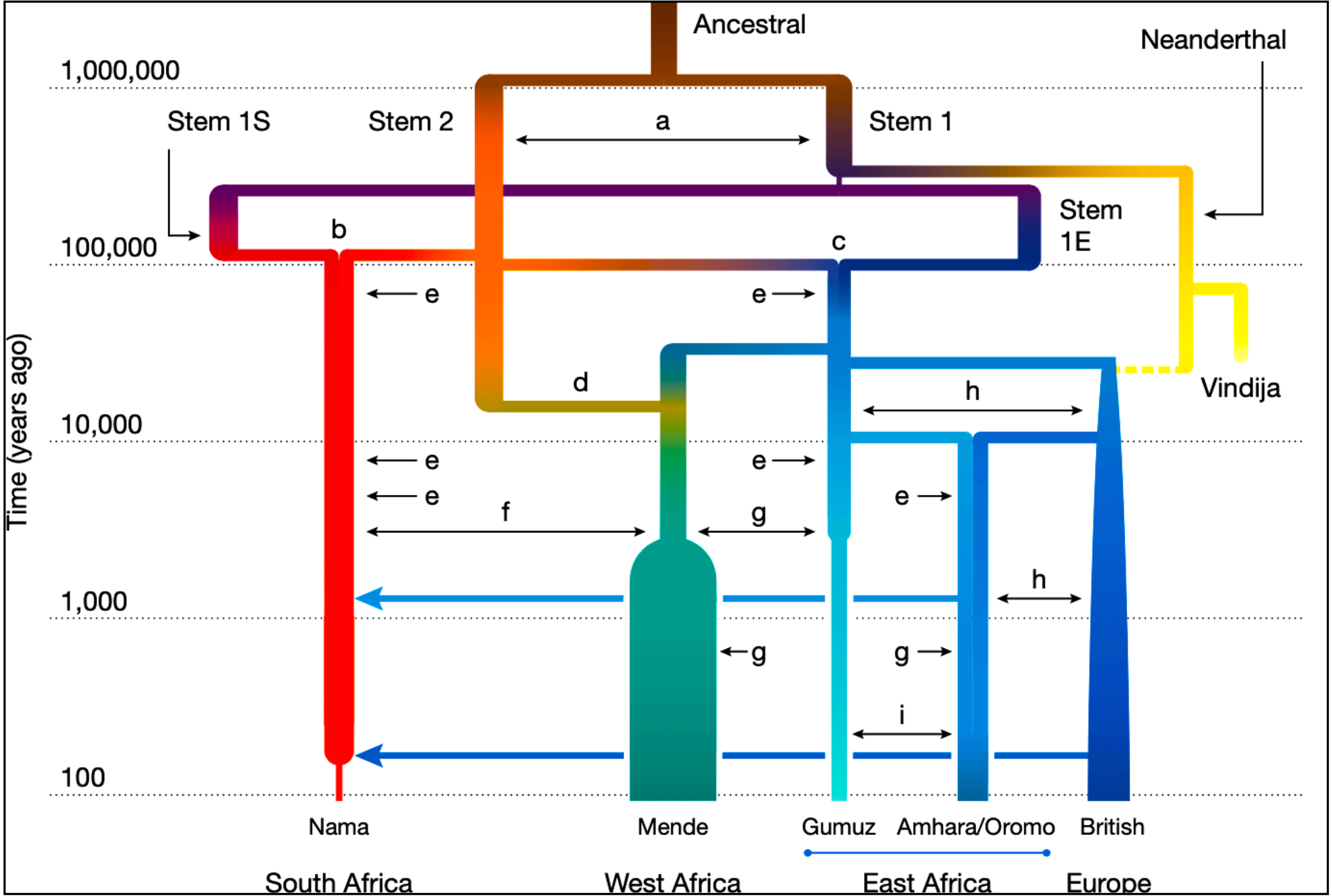
- Parameters $\Leftarrow$ population sizes, size changes, split times, continuous migration rates, admixture times, proportions
- Likelihoods $\Leftarrow$ Composite multivariate Gaussian likelihood approach
- Confidence intervals $\Leftarrow$ Refitting each model to 200 block-bootstrap replicate datasets

Methods

Gene genealogy reconstruction

- Reconstruct genome-wide gene genealogies → Software used: **Relate**
- *Coalescence rates* within and between populations → **effective population sizes**
- Compare reconstructed genealogies from data to model predictions to simulate genomic data for each population in models → Software used: **msprime**

A weakly structured stem best describes two-locus statistics



*The *weakly structured stem model* is **not incompatible** with *archaic admixture* having occurred in the ancestry of these fossils, but would imply, by contrast, that such individuals are **unlikely** to have contributed much ancestry to contemporary humans.

Conclusions

Highlights

1. Inferring a reticulated African population history in which present-day population structure dates back to Marine Isotope Stage 5 ~ **130,000 - 80,000 years ago**.
2. The model explains patterns of polymorphism that had previously been attributed to contributions from archaic hominins in Africa.
3. Fossil remains from coexisting ancestral populations should be genetically and morphologically similar (1-4% differentiation for genetic drift).
4. Model misspecification explains the variation in previous estimates of divergence times.

Takeaways

Highlights

- Geographical patterns of contemporary *H. sapiens* population structure probably arose during Marine Isotope Stage 5.
- Studying a range of models is key to making robust inferences about deep history.

Challenge: co-estimation of divergence time and subsequent migration

Future prospect: predictions of model to be related to the fossil record

“The knowledge that has not come down to us is
larger than the knowledge that has.”

Ibn Khaldun, The Muqaddimah (1377 AD)

Thanks for listening!