Ancient human extinction risk: Implications from high-precision computation and fossil evidence

Sumei Zhou[1,2]†, Yiwen Zhou[1]†, Fabio Di Vincenzo[3], Giorgio Manzi[3], Wangjie Hu[4], Ziqian Hao[5], Yi-Hsuan Pan[2]*, Haipeng Li[1]*

Affiliations:

[1]CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences; Shanghai, China.

[2]Key Laboratory of Brain Functional Genomics of Ministry of Education, School of Life Science, East China Normal University; Shanghai, China.

[3]Department of Environmental Biology, Sapienza University of Rome; Rome, Italy.

[4]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai; New York, USA.

[5]College of Medical Information and Artificial Intelligence, Shandong First Medical University & Shandong Academy of Medical Sciences; Jinan, China.

†These authors contributed equally to this work.
*Corresponding authors. Email: lihaipeng@sinh.ac.cn; yxpan@sat.ecnu.edu.cn

Abstract

Reconstructing ancient population size dynamics is fundamental to understanding human evolution. Using the fast infinitesimal time coalescent process (FitCoal), we identified a severe population bottleneck around 930,000 years ago that drastically reduced the ancestral population. Despite its profound evolutionary consequences, this event left only subtle genomic traces in present-day humans. Detecting this signal requires high-precision computation of expected branch lengths across site frequency spectrum categories. Our analyses demonstrate that FitCoal enables efficient and accurate demographic reconstructions, providing greater computational efficiency than alternative methods. Independent fossil and paleoclimate records further corroborate the timing and impact of this bottleneck, underscoring the necessity of integrating genetic, paleontological, and environmental evidence. These findings refine the understanding of hominin evolution and the demographic events that shaped modern human ancestry.

Keywords: Demographic inference; site frequency spectrum; FitCoal; Early-Middle Pleistocene

Over the past four million years, at least 22 hominin taxa have emerged (*1*), yet only modern humans, *Homo sapiens,* have survived to the present day. It is estimated that modern humans, Neanderthals, and Denisovans diverged from a common ancestor approximately 600 thousand years ago (Ka) (*2*). This ancestor likely arose as a derived species in Africa during the Middle Pleistocene transition, a period marked by the decline of African *H. erectus* (*3*). To reconstruct the evolutionary trajectory leading to modern humans, it is therefore essential to investigate their population size history from approximately 1,000 to 100 Ka, a timeframe coinciding with significant adaptive phenotypic changes.

We developed the fast infinitesimal time coalescent process (FitCoal) to infer both recent and ancient population size histories (*4*). FitCoal enables precise computation of the expected site frequency spectrum (SFS) and its associated likelihood under arbitrary demographic models. While a given SFS can theoretically be explained by different synthetic demographic models (*5*), FitCoal overcomes this challenge by prioritizing biologically realistic scenarios (*4, 6, 7*) and leveraging high-precision computation. To rigorously assess its accuracy in estimating population size histories, we tested FitCoal across 89 different models. Applying FitCoal to present-day human genomic sequences from the 1000 Genomes Project (*8*) and the Human Genome Diversity Project – Centre d'Etude du Polymorphisme Humain panel (*9*), we identified a severe population bottleneck occurring around 930 Ka in 12 African and 19 non-African populations (*4*). Notably, the timing of this bottleneck is nearly identical in both African and non-African populations (Figure 1A, B). More, we observed greater variance in bottleneck estimates in non-African populations than in African populations, likely due to the genetic signal dilution caused by out-of-Africa dispersal. Multiple lines of cross-disciplinary evidence further support the existence of this bottleneck, including an African hominin fossil gap (*10*), the fusion of two ancestral chromosomes (*11, 12*), significant cultural and behavioral shifts (*13-16*), climatic changes during the Early-to-Middle Pleistocene transition (*17-19*), and a corresponding increase in mammalian speciation and extinction rates (*20, 21*). Additionally, this severe bottleneck may contribute to a major human diaspora from Africa to Eurasia (*14, 22, 23*). Collectively, these findings underscore the necessity of integrating genetic, paleontological, and environmental evidence to substantiate the presence of this bottleneck.

To trace the human evolutionary trajectory, it is crucial to minimize confounding factors. High-frequency mutations are particularly susceptible to confounding influences, such as positive selection, introgression, population structure, and misclassification of derived and ancestral variants. FitCoal mitigates these effects by utilizing a less-impacted subset of the SFS or a collapsed SFS (*6, 7*), while maintaining high efficiency in inferring population size history. A key strength of FitCoal lies in its ability to extract population size history information not only from individual SFS categories but also from their relative ratios (*4*). To demonstrate this, we compared FitCoal-inferred histories using both the full SFS and a subset of the SFS. Remarkably, a bottleneck model was precisely inferred even when only 10% of SFS categories were analyzed (Figure 1C). These findings highlight the

multidimensional richness of demographic information embedded within the SFS, allowing FitCoal to remain robust despite the complexities of evolutionary history.

High-precision computation is critical for inferring ancient population size history. To evaluate FitCoal's accuracy, we tested a synthetic model in which FitCoal might falsely identify a bottleneck (*24*). First, we calculated FitCoal-expected branch length for each SFS category and compared it with results obtained using the original method (*24*). The maximum difference for each SFS category reached 6.21%, with a total discrepancy of 9,981 generations (Figure 1D, Table 1). To further assess FitCoal's computational precision, we conducted 1×10^{10} msprime simulations. The average branch lengths from these simulations were nearly identical to those calculated by FitCoal, reducing the total difference to only 21 generations. However, the required computing time differed substantially (0.49 seconds *vs.* 47 months), reflecting the computational efficiency of FitCoal. Additionally, we simulated 1,000 polymorphic datasets using msprime (*25*) under the synthetic model. In all cases, FitCoal correctly inferred the synthetic model (Figure 1E). These results highlight the pivotal role of simulated datasets in validating the ancient severe bottleneck and demonstrate that high-precision computation is essential for accurately computing the expected SFS and the SFS-related likelihood.

A comprehensive understanding of hominin evolution requires the integration of genetic, fossil, and paleoclimate records. Between 1,000 and 100 Ka, at least two major macroevolutionary events occurred (Figure 1F), coinciding with an African hominin fossil gap and the so-called 'muddle in the middle'— a period of hominin coexistence and morphological diversity (*26*). Given that small population size can drive speciation (*27*), our findings suggest that the severe bottleneck may have played a pivotal role in shaping phenotypic and behavioral evolution in human ancestors. In particular, the potential influence of the bottleneck on adaptation and speciation warrants further investigation. The fixation of fused chromosome 2 and its potential role in reproductive isolation remain key questions in human evolution. To explore these issues, we analyzed endocranial volume and the encephalization quotient in African fossil specimens. Brain size increased significantly following the severe bottleneck (Figure 1G, Table 2), suggesting a potential link between population contraction and accelerated cognitive and social evolution. Overall, these cross-disciplinary studies shed new light on hominin speciation, the genetic consequences of a severe bottleneck, and the evolutionary forces shaping modern human ancestry, refining our understanding of key events preceding the out-of-Africa dispersal.

Methods

To rigorously assess the accuracy of FitCoal in demographic inference, we tested it against a synthetic model from Deng et al. (*24*), used here as the true model. This model was derived by inferring demographic parameters while restricting the occurrence of a severe bottleneck, using a self-called expected SFS from a severe bottleneck model (*24*). The synthetic model is also similar to two previously investigated models (see Figure 4B and 4F in Hu et al. 2023). We obtained the synthetic model and Deng et al.'s expected branch lengths using their scripts. The expected SFS was calculated by multiplying the branch lengths by the total mutation rate of the considered sequence fragments.

We compared expected branch length across SFS category under the true model using three independent methods. First, Deng et al.'s expected branch lengths were obtained by dividing their expected SFS by the sequence length and the mutation rate per site per generation. Second, FitCoal's expected branch lengths were calculated using FitCoal version 1.3, which can be freely downloaded (https://doi.org/10.5281/zenodo.4765446). This version is released alongside this manuscript. Third, we used msprime (*25*) to simulate 1×10^{10} coalescent trees of a single locus based on the synthetic model. The mean branch length for each SFS category was then obtained. The number of simulations was determined by examining the accuracy of the expected branch lengths (Figure 2). Given the rounding of decimal values (Table 1), 1×10^{10} coalescent simulations were determined as the minimum required to achieve sufficient accuracy in inferring ancient population size history.

To examine the robustness of FitCoal in inferring population demography, we obtained 1,000 simulated SFSs under the synthetic model using msprime (*25*). The parameters of the simulation were aligned with a previously study (*24*), in which mutation and recombination rates were both set as 1.2×10^{-8} per base per generation with 216 sequences, each 100 Mb in length. Since the simulated data were generated under the single-population neutral model, there was no need to exclude high-frequency mutations. We used FitCoal with default parameters and inferred 1,000 population size histories.

The endocranial volume of Buia was obtained from Bruner et al. (*28*), KNM-ER 42700 from Neubauer et al. (*29*), and others from Holloway, Broadfield and Yuan (*30*). The encephalization quotient for each specimen was calculated following a previous study (*31*). Body mass (in kilograms) was obtained from three studies (*32-34*) to derive the quotient.

Acknowledgments: This work was supported by two grants of National Natural Science Foundation of China to YHP (no. 31100273) and to HL (no. 32270674).

Competing interests: Authors declare that they have no competing interests.

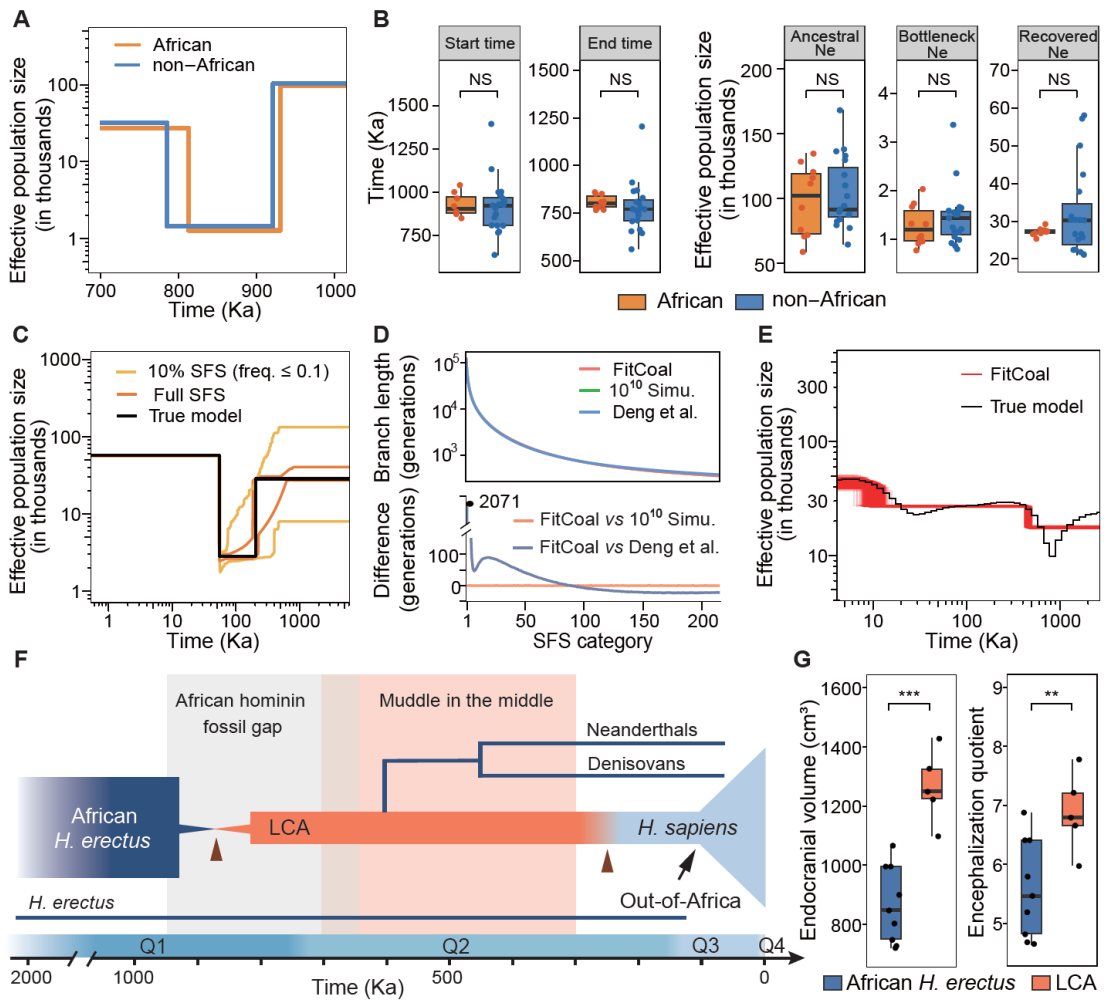

Figure 1. Implications of ancient human extinction risk, high-precision computation, and fossil evidence. (**A**) The severe bottleneck in African and non-African populations. (**B**) Estimated bottleneck parameters in 10 African and 19 non-African populations. Two South African populations (San and Bantu) were excluded due to their small sample sizes. (**C**) The 2.5th and 97.5th percentiles of FitCoal-inferred population size histories using Full-SFS and minor alleles ($freq \leq 0.1$). (**D**) The upper panel shows the expected branch length for each SFS category calculated by FitCoal, 1×10^{10} msprime simulations, and Deng et al. The lower panel displays their differences from FitCoal estimates. (**E**) FitCoal-inferred population size histories based on different SFSs under the synthetic model proposed by Deng et al. Red lines represent FitCoal-inferred histories based on 1,000 simulated datasets (n=216, L=100 Mb). (**F**) Evolutionary history of hominins over the past one million years. The LCA refers to the last common ancestor shared by Denisovans, Neanderthals, and modern humans. The two triangles indicate the speciation events. Q1 – Q4 indicate the Early Pleistocene, Middle Pleistocene, Late Pleistocene, and Holocene, respectively. (**G**) Endocranial volume and the encephalization quotient of African fossil specimens before and after the severe bottleneck. Two-tailed Student's t-test, ***P < 0.001, **P < 0.01 and NS = not significant. (**A – C**) Data were obtained from the previous results (*4*).

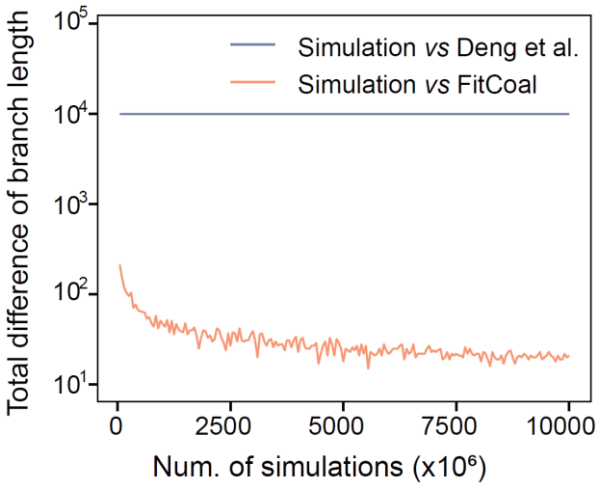

Figure 2. Influence of the number of simulations on the accuracy of averaged branch lengths compared with FitCoal and Deng et al. The synthetic model (Figure 1E) was used.

Table 1. Expected branch lengths (in generations) for each SFS category.

SFS category[a]	FitCoal (Generations)	1×10^{10} simulations (Generations)	Deng et al. (Generations)	Diff. FitCoal *vs.* simulations	Diff. FitCoal *vs.* Deng et al.
1	129,059	129,059	126,988	0 (0.00%)	2,071 (1.60%)
20	4,778	4,778	4,689	0 (0.00%)	89 (1.86%)
40	2,124	2,124	2,066	0 (0.00%)	58 (2.73%)
60	1,293	1,293	1,265	0 (0.00%)	28 (2.17%)
80	914	914	908	0 (0.00%)	6 (0.66%)
100	709	709	716	0 (0.00%)	7 (0.99%)
140	504	504	523	0 (0.00%)	19 (3.77%)
180	406	406	428	0 (0.00%)	22 (5.42%)
200	374	374	396	0 (0.00%)	22 (5.88%)
215	354	354	376	0 (0.00%)	22 (6.21%)
Total diff.[b]	–	–	–	21	9,981
Computing time	0.49 sec	47 Mon	8.75 sec	–	–

[a]The number of sampled sequences is 216, following Deng et al. (*24*).

[b]The summation of differences over all categories, including ones not shown in the table.

Table 2. Endocranial volume and the encephalization quotient of African *Homo erectus* and the LCA.

Taxon	Specimen	Endocranial volume (cm^3)	Encephalization quotient
H. erectus (Africa)	KNM-ER 3732	750	4.83
	KNM-ER 3733	848	5.46
	KNM-ER 3883	804	5.18
	Daka	995	6.41
	Buia	995	6.41
	KNM-ER 42700	721	4.64
	KNM-WT 15000	900	5.79
	OH 9	1,067	6.87
	OH 12	727	4.68
LCA	Bodo	1,250	6.80
	Broken Hill 1	1,325	7.21
	Saldanha	1,225	6.66
	Omo-Kibish 2	1,430	7.77
	Ndutu	1,100	5.98

References

1. A. A. de Sousa *et al.*, From fossils to mind. *Commun Biol* **6**, 636 (2023).
2. K. Prüfer *et al.*, A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* **358**, 655-658 (2017).
3. F. Di Vincenzo, G. Manzi, *Homo heidelbergensis* as the Middle Pleistocene common ancestor of Denisovans, Neanderthals and modern humans. *Journal of Mediterranean Earth Sciences* **15**, 303-315 (2023).
4. W. Hu *et al.*, Genomic inference of a severe human bottleneck during the Early to Middle Pleistocene transition. *Science* **381**, 979-984 (2023).
5. S. Myers, C. Fefferman, N. Patterson, Can one learn history from the allelic spectrum? *Theor Popul Biol* **73**, 342-348 (2008).
6. X. M. Liu, Y. X. Fu, Exploring population size changes using SNP frequency spectra. *Nat Genet* **47**, 555-559 (2015).
7. A. Bhaskar, Y. X. Rachel Wang, Y. S. Song, Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *Genome Res* **25**, 268-279 (2015).
8. The 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
9. A. Bergström *et al.*, Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
10. A. Profico, F. Di Vincenzo, L. Gagliardi, M. Piperno, G. Manzi, Filling the gap. Human cranial remains from Gombore II (Melka Kunture, Ethiopia; ca. 850 ka) and the origin of *Homo heidelbergensis*. *J Anthropol Sci* **94**, 41-63 (2016).
11. B. Poszewiecka, K. Gogolewski, P. Stankiewicz, A. Gambin, Revised time estimation of the ancestral human chromosome 2 fusion. *BMC Genomics* **23**, 616 (2022).
12. T. R. Dreszer, G. D. Wall, D. Haussler, K. S. Pollard, Biased clustered substitutions in the human genome: the footprints of male-driven biased gene conversion. *Genome Res* **17**, 1420-1430 (2007).
13. J. Clark, G. J. Linares-Matás, When to generalise and when to specialise? Climate change and hominin biocultural adaptability in the African early and middle stone age. *Quat Sci Adv* **15**, 100218 (2024).
14. M. H. Moncel *et al.*, The origin of early Acheulean expansion in Europe 700 ka ago: new findings at Notarchirico (Italy). *Sci Rep* **10**, 13802 (2020).
15. D. M. Sandgathe, Identifying and describing pattern and process in the evolution of hominin use of fire. *Current Anthropology* **58**, S360-S370 (2017).
16. R. Gallotti, M. Mussi, Two Acheuleans, two humankinds: From 1.5 to 0.85 ma at Melka Kunture (upper awash, Ethiopian highlands). *J Anthropol Sci* **95**, 137-181 (2017).
17. J. Jouzel *et al.*, Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317**, 793-796 (2007).
18. M. J. Head, P. L. Gibbard, in *Early-Middle Pleistocene Transitions: The Land–Ocean Evidence*, M. J. Head, P. L. Gibbard, Eds. (Geological Society of London, London, 2005), vol. 247, pp. 1-18.
19. P. U. Clark *et al.*, The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO_2. *Quat Sci Rev* **25**, 3150-3184

(2006).
20. A. S. Cohen *et al.*, Plio-Pleistocene environmental variability in Africa and its implications for mammalian evolution. *Proc Natl Acad Sci USA* **119**, e2107393119 (2022).
21. A. Timmermann *et al.*, Past climate change effects on human evolution. *Nature Reviews Earth & Environment* **5**, 701-716 (2024).
22. G. Muttoni, D. V. Kent, Hominin population bottleneck coincided with migration from Africa during the Early Pleistocene ice age transition. *Proc Natl Acad Sci USA* **121**, e2318903121 (2024).
23. I. Micarelli *et al.*, The oldest fossil hominin from Italy: Reassessment of the femoral diaphysis from Venosa-Notarchirico in its Acheulean context. *Quat Sci Rev* **334**, 108709 (2024).
24. Y. Deng, R. Nielsen, Y. S. Song, A previously reported bottleneck in human ancestry 900 kya is likely a statistical artifact. *Genetics*, iyae192 (2024).
25. F. Baumdicker *et al.*, Efficient ancestry and mutation simulation with msprime 1.0. *Genetics* **220**, iyab229 (2022).
26. A. Bergström, C. Stringer, M. Hajdinjak, E. M. L. Scerri, P. Skoglund, Origins of modern human ancestry. *Nature* **590**, 229-237 (2021).
27. S. Gould, N. Eldredge, Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology* **3**, 115-151 (1977).
28. E. Bruner *et al.*, The endocast of the one-million-year-old human cranium from Buia (UA 31), Danakil Eritrea. *Am J Phys Anthropol* **160**, 458-468 (2016).
29. S. Neubauer *et al.*, Reconstruction, endocranial form and taxonomic affinity of the early *Homo* calvaria KNM-ER 42700. *J Hum Evol* **121**, 25-39 (2018).
30. R. L. Holloway, D. C. Broadfield, M. S. Yuan, *The Human Fossil Record: Brain Endocasts - The Paleoneurological Evidence*. J. H. Schwartz, I. Tattersall, Eds., The Human Fossil Record (John Wiley & Sons, Inc., New Jersey, 2004), vol. 3.
31. H. J. Jerison, *Evolution of the Brain and Intelligence*. (Academic Press, New York, 1973).
32. M. Grabowski, K. G. Hatala, W. L. Jungers, B. G. Richmond, Body mass estimates of hominin fossils and the evolution of human body size. *J Hum Evol* **85**, 75-93 (2015).
33. W. L. Jungers, M. Grabowski, K. G. Hatala, B. G. Richmond, The evolution of body size and shape in the human career. *Philos Trans R Soc Lond B Biol Sci* **371**, 20150247 (2016).
34. J. M. Carretero *et al.*, *Los humanos de la Sima de los Huesos (Sierra de Atapuerca) y la evolución del cuerpo en el género Homo.* E. Baquedano, S. Rubio, Eds., Miscelánea en homenaje a Emiliano Aguirre (Museo arqueológico regional, Alcalá de Henares, Spain, 2004), vol. 3.