Genomic inference of a human super bottleneck in Mid-Pleistocene transition

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

[1]CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences; Shanghai 200031, China.

[2]Key Laboratory of Brain Functional Genomics of Ministry of Education, School of Life Science, East China Normal University; Shanghai 200062, China.

[3]Natural History Museum, University of Florence; Florence, Italy.

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

[5]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences; Kunming 650223, China.

*Corresponding Authors: yxpan@sat.ecnu.edu.cn; lihaipeng@picb.ac.cn

†These authors contributed equally to this work.

Abstract: The demographic history is a foundation of human evolutionary studies. We developed the fast infinitesimal time coalescent (FitCoal) process, which allows the accurate calculation of the composite likelihood of a site frequency spectrum and provides the precise inference of demographic history. Genomic analysis showed that African populations have passed through a population super bottleneck, a small effective size of approximately 1,280 breeding individuals between 930 and 813 thousand years ago. This time interval coincides with a gap in the human fossil record in Africa and possibly marks the origin of the *Homo heidelbergensis*. Further modelling analysis confirmed the existence of the super bottleneck in non-African populations. Our results provide new insights into human evolution during the Mid-Pleistocene.

One-Sentence Summary: A new method for demographic history inference and a human super bottleneck possibly marking the origin of *H. heidelbergensis*

Inferring demographic history from genomic information has played an important role in population genetics. It uncovers prehistoric evolutionary events and deepens our understanding about the evolution of human and other species (*1-3*). Multiple methods have been developed to infer demographic history with a predefined demographic model (*4-7*). These methods require prior knowledge about the species being investigated and estimation parameters by fitting in summary statistics such as site frequency spectrum (SFS). In contrast, model-free methods do not need a predefined model when inferring demography (*8-13*). As SFS plays an essential role in demographic inference, many efforts have been attempted to derive its analytical formula under certain demographic models (*14-16*).

To precisely infer recent and ancient demography, we developed the fast infinitesimal time coalescent (FitCoal) process (Fig. 1) that calculates expected branch length for each SFS type under arbitrary demographic models. It is effective for a wide range of sample sizes in the calculation of the composite likelihood of a given SFS (*4, 5*). To infer the demographic history, FitCoal first maximizes the likelihood with the constant size model and then increases the number of inference time intervals and re-maximizes the likelihood until the best model is found. As inference time intervals are variable to avoid long fixed ancient time intervals, the precision in the inference of ancient demographic events can be improved. FitCoal does not need prior information on demography, and its accuracy can be confirmed by simulation.

With African hominid fossils, the origin of anatomically modern humans has been determined to be approximately 200 thousand years (kyr) ago (*17*). Although the demographic history of humans has been intensively studied (*9, 10, 12, 18-21*), it is conceivable that many new insights remain to be explored. Here we used FitCoal to analyze genomic sequences of the 1000 Genomes Project phase 3 (1000GP) (*22*) and the Human Genome Diversity Project–Centre d'Etude du Polymorphisme Humain panel (HGPD-CEPH) (*23*). Our results suggest that our ancestors experienced a super bottleneck and the effective size of our ancestors remained small for a very long

period of time (~117,000 years). The super bottleneck may represent a major transition in human evolution, possibly leading to the origin of *H. heidelbergensis*: the alleged ancestral species of modern humans (*24*, *25*).

Fast infinitesimal time coalescent process

As determination of expected branch length for each SFS type is essential for theoretical population genetics and demographic inference (*4*, *5*), we developed the fast infinitesimal time coalescent (FitCoal) process to accomplish the task (Fig. 1). For FitCoal analysis, each of millions of time intervals Δt was set extremely small, and the population size was assumed to be constant within each infinitesimal time interval. The probabilities of all states were calculated backward in time. During each Δt, the branches were categorized according to their state. For each state, the branch length was multiplied by its probability and population size and then transformed to calculate the expected branch length of each SFS type. Because the expected branch length of a SFS type is equal to the sum of the expected branch length of this type during each time interval, the latter can be rescaled and tabulated, making the calculation of the expected branch lengths extremely fast under arbitrary demographic histories. Hereafter, tabulated FitCoal is referred to as FitCoal for short, unless otherwise indicated.

FitCoal demographic inference

After the expected branch lengths were obtained, the composite likelihood of the SFS observed in a sample was calculated (*4*, *5*, *10*, *26*). As each single nucleotide polymorphism (SNP) was treated independently, FitCoal did not need phased haplotype data. When inferring demography, the likelihood was maximized in a wide range of demographic scenarios. The FitCoal likelihood surface is smooth (Fig. S1), so it is efficient to maximize the likelihood. FitCoal considered both instantaneous populations size changes (*9*-*11*) and long-term exponential changes of population in order to generate various demographic scenarios.

Demographic inference on simulated data

The accuracy of FitCoal was validated by simulation and comparing its demographic inferences with those of PSMC (*9*) and stairway plot (*10*) (Fig. 2). Six different demographic models, examined in the study by Liu and Fu (*10*), were considered by simulating 200 independent data sets under each model. The medians and 95% confidence intervals of demography were then determined by FitCoal with the assumption that a generation time is 24 years (*10, 27*) and the mutation rate is 1.2×10^{-8} per site per generation for human populations (*10, 28-30*).

FitCoal was found to precisely infer demographic histories (Fig. 2), and the inference accuracy was improved by increasing sample size and length of sequence (Fig. S2). Our results confirmed that SFS allows precise recovery of the demographic history (*31*). In general, the confidence intervals of FitCoal inferred histories were narrower than those of PSMC and stairway plot, except for those with insufficient information on ancestral populations (Fig. 2C). The proportion of the most recent change type inferred from the six different models mentioned above also showed that FitCoal can distinguish instantaneous and exponential changes (Table S1).

Since a demographic event may affect every SFS type, demographic history can be inferred using a subset of SFS. Results of simulation confirmed that FitCoal accurately determined demographic history based on truncated SFSs (Fig. S3, S4), thus reducing the impact of other factors, such as positive selection (Fig. S5) and sequencing error, on FitCoal analysis.

Demographic inference of African populations

To infer the demographic histories of African populations, seven African populations in the 1000GP (*22*) were analyzed by FitCoal. Only non-coding regions, defined by GENCODE (*32*), were used in order to avoid the effect of purifying selection. To avoid the potential effect of positive selection (*33*), high-frequency mutations were excluded from the analysis.

Results showed that all seven African populations passed through a super bottleneck around 914 (854–1,003) kyr ago and that this bottleneck was relieved about 793 (772–815) kyr ago (Fig. 3A-C, S6; Table S2). The average effective population size of African populations during the bottleneck period was determined to be 1,270 (770–2,030). Although traces of the bottleneck were observed in previous studies, the bottleneck was ignored because its signatures were too weak to be noticed

(*9*, *11*, *12*, *22*, *23*). After the bottleneck was relieved, the population size was increased to 27,080 (25,300–29,180), a 20-fold increase, around 800 kyr ago. This population size remained relatively constant until the recent expansion.

To avoid the potential effects of low sequencing depth (~ 5x) of non-coding regions in the 1000GP on the analysis, the autosomal non-coding genomic polymorphism of HGDP-CEPH data set with high sequencing coverage (~35x) was used. In total, populations with more than 15 individuals each were examined. Results showed that the super bottleneck occurred between 1,257 (1,042–1,527) and 859 (856–864) kyr ago in all three African populations in HGDP-CEPH (Fig. 3D-F, S7; Table S3), and the average population size during the bottleneck period was 1,300 (908–1,670). This number was very similar to that (1,270) estimated from the data of 1000GP.

After the bottleneck was relieved, the population sizes of the two HGDP-CEPH agriculturalist populations were increased to 27,300 and 27,570 (Fig. 3E, S7; Table S3), consistent with the 1000GP estimate of 27,280. However, the Biaka, a hunter-gatherer population, had a larger population size of 35,330, suggesting a deep divergence between this and other agriculturalist populations (*34*-*36*). The Biaka population was found to have a recent population decline (Fig. 3D, S7), as previously observed (*23*). These results suggest that hunter-gatherer populations were widely spread and decreased when agriculturalist populations were expanded.

To provide a precise inference of the super bottleneck, the results from the two data sets were combined. After analyzing the inferred time of instantaneous change of 10 populations, the super bottleneck was inferred to last for about 117,000 years, from 930 (854–1,042; s.e.m.: 23.52) to 813 (772–864; s.e.m.: 11.02) kyr ago. The effective size during the bottleneck period was precisely determined to be 1,280 (767–2,031; s.e.m.: 131). A loss of 65.85% in current genetic diversity of human populations was estimated because of the bottleneck.

Demographic inference of non-African populations

No super bottleneck was directly observed in all 19 non-African populations in 1000GP (Fig. 3A-C, S6; Table S4). The ancestral population size of these populations was determined to be 20,260 (18,850–22,220), similar to that determined in previous studies (*9*, *11*, *12*, *23*). The population size started to decline around 368 (175–756)

kyr ago in 1000GP non-African populations, suggesting that African and non-African divergence occurred much earlier than the out-of-Africa migration (*9, 11, 12, 19, 22, 23*). European and South Asian populations were found to have a relatively weaker bottleneck than East Asian populations, and the bottleneck severity was found to correlate with their geographic distance to Africa, consistent with the observed correlation between heterozygosity and geographic distance (*21, 37*). A weak bottleneck was observed in American populations, probably because of recent admixture (*22*). All 1000GP non-African populations were found to increase in size recently.

The super bottleneck was also not directly detected in all 21 HGDP-CEPH non-African populations (Fig. 3D-F S7; Table S5). The ancestral population size of these populations was determined to be 20,030 (19,060–21,850), very similar to that (20,260) estimated from 1000GP. These populations started to decline 367 (167–628) kyr ago. A positive correlation was also observed between the severity of out-of-Africa bottleneck and their geographic distance to Africa. The Middle East populations had the weakest bottleneck, while the Maya, an American population, had the strongest bottleneck. Similar to 1000GP non-African populations, most HGDP-CEPH non-African populations were found to increase in size recently, except an isolated Kalash population, consistent with previous studies (*23, 38*).

Super bottleneck in the early Middle Pleistocene

Although a super bottleneck was detected in all 10 African populations, such bottleneck was not directly detected in all 40 non-African populations. To investigate this phenomenon, simulations were performed with three 1000GP demographic models, designated Bottleneck I, II, and III (Fig. 4). Bottleneck I simulated the average inferred demographic history of African populations with the super bottleneck, and Bottleneck II and III simulated the demography of non-African populations without and with the super bottleneck. Both Bottleneck I and II were inferred correctly as a corresponding bottleneck was found in all simulated data sets (Table S6). However, no super bottleneck was detected in Bottleneck III simulations. The super bottleneck was found to cause a population size gap between the true model and inferred demographic history after the bottleneck was relieved, suggesting a hidden effect of the super bottleneck on non-African populations. Simulations were

then extended to HGDP-CEHP populations with Bottleneck models IV–VI, and similar results were obtained (Fig. S8; Table S7). When simulations were performed on three artificial models (Bottleneck VII–IX) with various demographic parameters, a population size gap was still detected (Fig. S9; Table S8). These results suggest a hidden effect of the super bottleneck on non-African populations.

The population size gap was found in both 1000GP and HGDP-CEPH data sets (Fig. 3A, D). After the bottleneck was relieved, the average population sizes of non-African populations were determined to be 20,260 and 20,030, respectively, while those of African agriculturalist populations were 27,080 and 27,440, respectively in these two data sets. The observed population size gap was 7,020, probably due to the hidden effect of the super bottleneck on non-African populations.

The reasons were then investigated why the super bottleneck had different effects on African and non-African populations. Results showed that non-African populations had the out-of-Africa bottleneck, but African populations lacked such bottleneck. Therefore, the standard coalescent time of non-African populations was larger than that of African populations (Fig. 3C, F). As African populations had more coalescent events occurred during the bottleneck period, the bottleneck was more readily detected. The mathematical proof on this issue was described in the supplemental material.

Discussion

In this study, we develop FitCoal, a novel model-flexible tool for demographic inference. Key characteristic features of FitCoal are that the composite likelihood can be rapidly calculated based on expected branch lengths and that inference time intervals are variable during the demographic inference. Since coalescent events become rare when tracing backward in time, the length of time interval is usually set to increase progressively (*9*-*12*). Although this strategy can capture recent demographic events, it may miss ancient ones. As FitCoal uses variable time intervals, it can give an accurate inference for both recent and ancient demographic events.

The most important discovery with FitCoal in this study is that human ancestors passed through a super bottleneck. The ancient population size reduction around 930 kyr ago was likely to be driven by the climatic changes at the transition between the Early and Middle Pleistocene (*39*). During the transition, low-amplitude 41 kyr

obliquity-dominated glacial cycles shifted to quasi-periodic, low frequency 100 kyr periodicity, and climate change became more extreme and unpredictably associated with a longer dry period in Africa and a large faunal turnover in Africa and Eurasia (*40*). Coinciding with this date, archaic humans referable to African *Homo erectus* became extinct. Subsequently, from about 900 until 600 kyr ago, there is a gap in the human fossil record in Africa (Fig. S10). Only few fossil specimens have been found in this time span, such as the cranial fragments from Gombore in Ethiopia and the mandibles from Tighenif in Algeria, all of which show features linked to later *H. heidelbergensis* representatives and represent the evolutionary origin of this species (*24*). As a matter of fact, our data suggest that the ancestors of modern humans had a very small effective size of approximately 1,280 breeding individuals during the bottleneck period. This number is comparable in the same magnitude in the effective size of mammals threatened by extinction (*41*).

A rapid population recovery was inferred in all 10 African populations with a 20-fold population growth during a short time period around 813 kyr ago. The earliest archaeological evidence for human control of fire was found in Israel 790 kyr ago (*42*). As the control of fire profoundly affected social evolution (*43*) and brain size (*44*), it may be associated with the big bang in population size at the end of the super bottleneck.

The super bottleneck, which started about one million years ago, might represent a speciation event at the origin of *H. heidelbergensis* (*24*) and should be strongly related to the gap in the fossil record. The questions about where the small ancient population dwelt, and how they survived for such a long time, remain to be investigated. As Neanderthals and Denisovans diverged with modern humans between 440 and 270 kyr ago (*45, 46*), it is conceivable that their ancestors also passed through the super bottleneck. In the future, a more detailed picture of human evolution during the Pleistocene may be revealed because more genomic sequences of present populations and those of archaic hominins as well as more advanced population genomics methods will be available.

Acknowledgement: We thank Daniel Zivković for sharing his codes to calculate the expected branch length, and Xiaoming Liu for sharing his simulated results. This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13040800), the National Natural Science Foundation of China (nos. 31100273, 31172073, 91131010), and National Key Research and Development Project (No. 2020YFC0847000).

Author contributions: W.H., Z.H., Y.H.P., and H.L. conceived and designed the research; W.H., Z.H., and H.L. wrote the code; W.H., Z.H., P.D., F.D.V., G.M., and Y.H.P. analyzed the data; W.H., Z.H., P.D., F.D.V., G.M., Y.H.P., and H.L. wrote the paper.

Competing interests: The authors declare no competing interests.

Data and materials availability: The authors declare that all data are available in the main text or the supplementary materials. FitCoal is a free plug-in of the eGPS software (*47*) and can be downloaded and run as an independent package. FitCoal and its documentation are available via Zenodo at https://zenodo.org/record/4765447#.YKDt7aG-vuq, our institute website at http://www.picb.ac.cn/evolgen/, and eGPS website http://www.egps-software.net/.

References and Notes

1. J. Y. Hu *et al.*, Genomic consequences of population decline in critically endangered pangolins and their demographic histories. *Natl. Sci. Rev.* **7**, 798-814 (2020).
2. L. Zeng *et al.*, Out of southern East Asia of the brown rat revealed by large-scale genome sequencing. *Mol. Biol. Evol.* **35**, 149-158 (2018).
3. X. Mao *et al.*, The deep population history of northern East Asia from the Late Pleistocene to the Holocene. *Cell*, (2021).
4. H. Li, W. Stephan, Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet.* **2**, e166 (2006).
5. L. Excoffier, I. Dupanloup, E. Huerta-Sanchez, V. C. Sousa, M. Foll, Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).
6. R. C. Griffiths, S. Tavaré, Monte Carlo inference methods in population genetics. *Math. Comput. Modell.* **23**, 141-158 (1996).
7. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
8. J. Heled, A. J. Drummond, Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* **8**, 289 (2008).
9. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496 (2011).
10. X. M. Liu, Y. X. Fu, Exploring population size changes using SNP frequency spectra. *Nat. Genet.* **47**, 555-559 (2015).
11. S. Schiffels, R. Durbin, Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919-925 (2014).
12. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303-309 (2017).
13. X. Liu, Y. X. Fu, Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* **21**, 280 (2020).
14. J. Jouganous, W. Long, A. P. Ragsdale, S. Gravel, Inferring the joint demographic history of

multiple populations: beyond the diffusion approximation. *Genetics* **206**, 1549-1567 (2017).

15. Y. X. Fu, Statistical properties of segregating sites. *Theor. Popul. Biol.* **48**, 172-197 (1995).

16. D. Zivković, T. Wiehe, Second-order moments of segregating sites under variable population size. *Genetics* **180**, 341-357 (2008).

17. T. D. White *et al.*, Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature* **423**, 742-747 (2003).

18. M. Stoneking, J. Krause, Learning about human population history from ancient and modern genomes. *Nat. Rev. Genet.* **12**, 603-614 (2011).

19. R. Nielsen *et al.*, Tracing the peopling of the world through genomics. *Nature* **541**, 302-310 (2017).

20. A. Manica, W. Amos, F. Balloux, T. Hanihara, The effect of ancient population bottlenecks on human phenotypic variation. *Nature* **448**, 346-348 (2007).

21. S. Ramachandran *et al.*, Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. USA* **102**, 15942-15947 (2005).

22. D. M. Altshuler *et al.*, A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

23. A. Bergstrom *et al.*, Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).

24. C. Stringer, The origin and evolution of Homo sapiens. *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.* **371**, 20150237 (2016).

25. 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).

26. R. R. Hudson, Two-locus sampling distributions and their application. *Genetics* **159**, 1805-1817 (2001).

27. A. Scally, R. Durbin, Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* **13**, 745-753 (2012).

28. C. D. Campbell *et al.*, Estimating the human mutation rate using autozygosity in a founder population. *Nat. Genet.* **44**, 1277-1281 (2012).

29. D. F. Conrad *et al.*, Variation in genome-wide mutation rates within and between human families. *Nat. Genet.* **43**, 712-714 (2011).

30. A. Kong *et al.*, Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471-475 (2012).

31. A. Bhaskar, Y. S. Song, Descartes' rule of signs and the identifiability of population demographic models from genomic variation data. *Ann. Stat.* **42**, 2469-2493 (2014).

32. A. Frankish *et al.*, GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766-D773 (2019).

33. J. C. Fay, C.-I. Wu, Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405-1413 (2000).

34. P. Hsieh *et al.*, Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies. *Genome Res.* **26**, 291-300 (2016).

35. C. M. Schlebusch, M. Jakobsson, Tales of human migration, admixture, and selection in Africa. *Annu. Rev. Genomics Hum. Genet.* **19**, 405-428 (2018).

36. P. Skoglund *et al.*, Reconstructing prehistoric African population structure. *Cell* **171**, 59-71 (2017).

37. F. Prugnolle, A. Manica, F. Balloux, Geography predicts neutral genetic diversity of human populations. *Curr. Biol.* **15**, R159-160 (2005).

38. Q. Ayub *et al.*, The Kalash genetic isolate: ancient divergence, drift, and selection. *Am. J. Hum. Genet.* **96**, 775-783 (2015).

39. L. E. Lisiecki, M. E. Raymo, A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography* **20**, PA1003 (2005).

40. M. J. Head, B. Pillans, S. A. Farquhar, The Early-Middle Pleistocene transition: characterization and proposed guide for the defining boundary. *Episodes* **31**, 255 (2008).

41. H. Li *et al.*, Large numbers of vertebrates began rapid population decline in the late 19th century. *Proc. Natl. Acad. Sci. USA* **113**, 14079-14084 (2016).

42. N. Goren-Inbar *et al.*, Evidence of hominin control of fire at Gesher Benot Ya'aqov, Israel. *Science* **304**, 725-727 (2004).

43. R. Foley, C. Gamble, The ecology of social transitions in human evolution. *Philos. Trans. R.*

Soc. Lond., Ser. B: Biol. Sci. **364**, 3267-3279 (2009).

44. M. Melchionna *et al.*, From smart apes to human brain boxes. A uniquely derived brain shape in late hominins clade. *Front. Earth Sci.* **8**, 273 (2020).

45. R. E. Green *et al.*, A draft sequence of the Neandertal genome. *Science* **328**, 710-722 (2010).

46. D. Reich *et al.*, Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060 (2010).

47. D. Yu *et al.*, eGPS 1.0: comprehensive software for multi-omic and evolutionary analyses. *Natl. Sci. Rev.* **6**, 867-869 (2019).

48. H. C. Harpending *et al.*, Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA* **95**, 1961-1967 (1998).

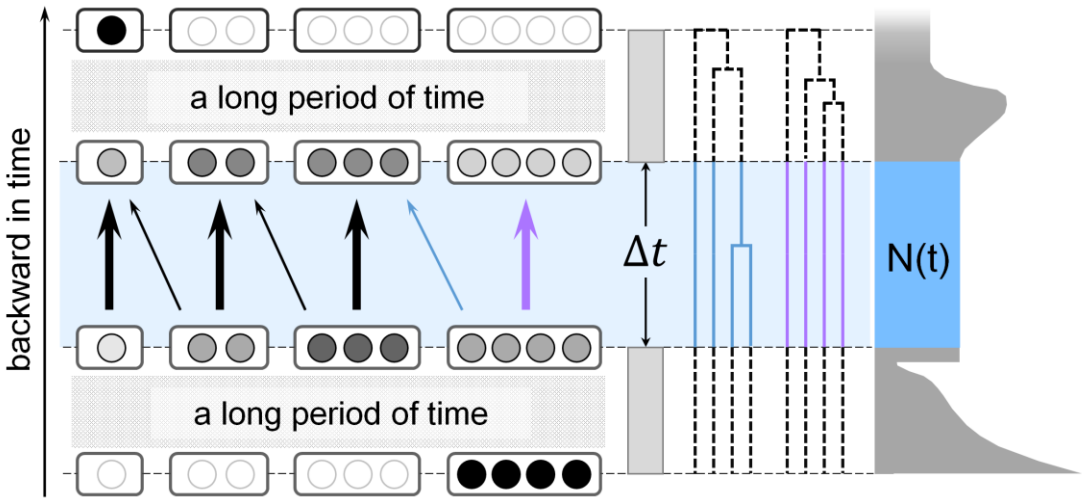


Fig. 1. Illustration of the fast infinitesimal time coalescent (FitCoal) process. The left panel shows the backward process in which four lineages coalesce into one after passing through millions of infinitesimal time intervals. The highlighted area shows the backward transformation process of different states with tiny probability changes in an infinitesimal time interval (Δt). Thick arrows indicate high transformation probabilities, and thin arrows indicate low transformation probabilities. Each state is indicated with a rounded rectangle, in which one circle indicates one lineage. The rounded rectangles with black filled circles are the states with probability 1. The rounded rectangles with empty circles are the states with probability 0. The probabilities between 0 and 1 are indicated by grey circles. The middle panel shows branches of different states. The right panel shows the demographic history of a population. The width of shadowed area indicate the effective population size, *i.e.*, the number of breeding individuals (*48*). It is assumed that the effective population size remains unchanged within Δt.

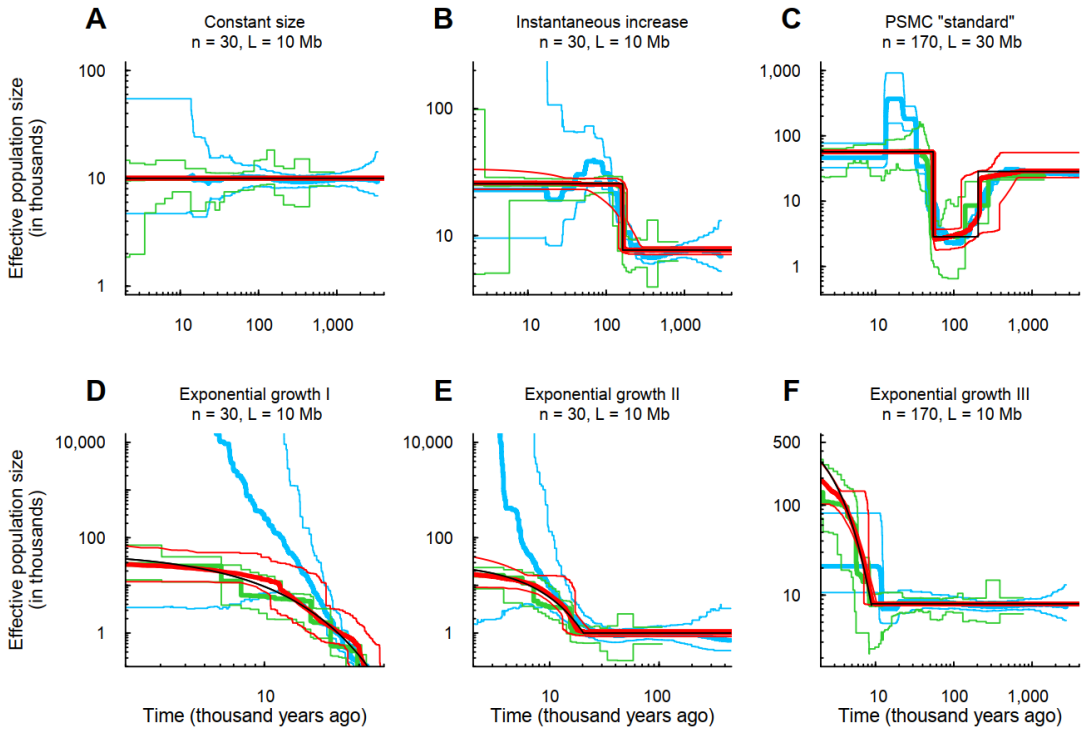


Fig. 2. Demographic histories estimated by FitCoal, stairway plot, and PSMC using simulated samples. (**A**) Constant size model. (**B**) Instantaneous increase model. (**C**) PSMC "standard" model. (**D**) Exponential growth I model. (**E**) Exponential growth II model. (**F**) Exponential growth III model. These six models are the same as those of the previous study by Liu and Fu (*10*). Thin black lines indicate true models. Thick red lines indicate the medians of FitCoal estimated histories; thin red lines are 2.5 and 97.5 percentiles of FitCoal estimated histories. Green and blue lines indicate the results of stairway plot and PSMC, respectively, of the previous study (*10*). The mutation rate is assumed to be 1.2×10^{-8} per base per generation, and a generation time is assumed to be 24 years. n is the number of simulated sequences, and L is the length of simulated sequences.

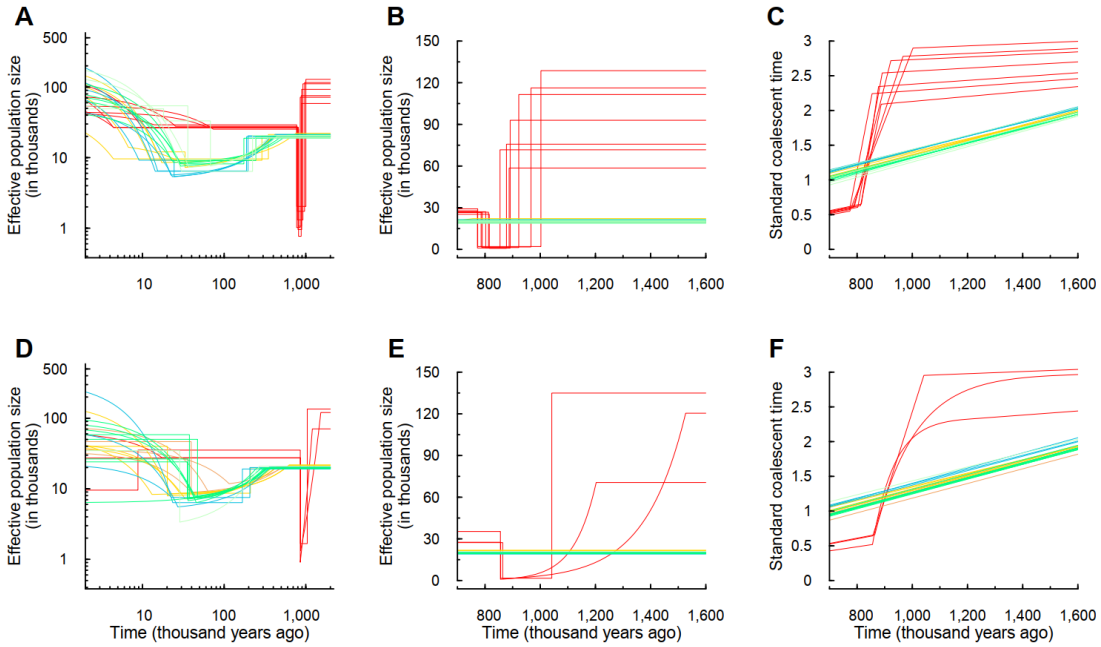
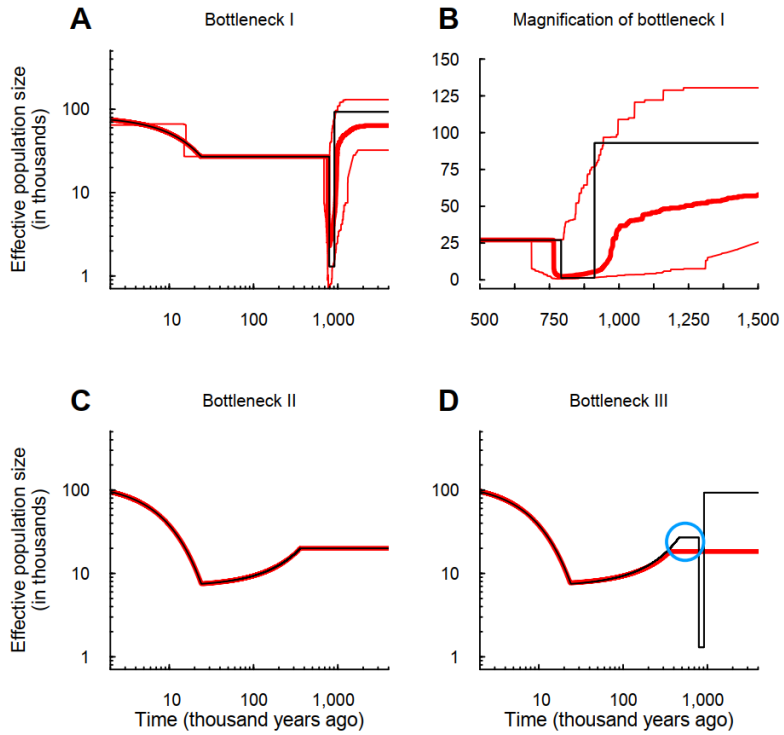


Fig. 3. FitCoal estimated histories of human populations using 1000GP and HGPD-CEPH genomic data sets. (**A**) Estimated histories of 26 populations in 1000GP. (**B**) Linear-scaled estimation of histories of 1000GP populations during the super bottleneck period. (**C**) Calendar time *vs* standard coalescent time of estimated histories of 1000GP populations. (**D**) Estimated histories of 24 HGPD-CEPH populations. (**E**) Linear-scaled estimation of histories of HGPD-CEPH populations during the super bottleneck period. (**F**) Calendar time *vs* standard coalescent time of estimated histories of HGPD-CEPH populations. Various color lines indicate the following: red, African populations; yellow, European populations; brown, Middle East populations; blue, East Asian populations; green, Central or South Asian populations; and dark sea green, American populations. The mutation rate is assumed to be 1.2×10^{-8} per base per generation, and a generation time is assumed to be 24 years.



426

427 **Fig. 4. Verification of the super bottleneck.** (**A**) Bottleneck I model, mimicking the
428 demography of 1000GP African population and its estimated histories. (**B**)
429 Linear-scaled Bottleneck I model during the super bottleneck period. (**C**) Bottleneck
430 II model, mimicking the estimated demography of 1000GP non-African population
431 and its estimated histories. (**D**) Bottleneck III model, mimicking the true demography
432 of 1000GP non-African population and its estimated histories. Thin black lines
433 indicate models. Thick red lines denote the medians of FitCoal estimated histories;
434 thin red lines represent 2.5 and 97.5 percentiles of FitCoal estimated histories. Blue
435 circle indicates the population size gap. The mutation rate is assumed to be
436 1.2×10^{-8} per base per generation, and a generation time is assumed to be 24 years.
437 The number of simulated sequences is 202 in Bottleneck I and 200 in Bottleneck II
438 and III. The length of simulated sequence is 800 Mb.

439

440

441

442