



The uncertainty of population relationship and divergence time inferred by the multiple sequentially Markovian coalescent model

Xing-Yan Yang^{1,2} · Shan-Shan Dai^{2,3,4} · He-Qun Liu² · Min-Sheng Peng^{2,3,4}  · Ya-Ping Zhang^{1,2,3,4}

Received: 30 October 2017 / Revised: 16 January 2018 / Accepted: 25 January 2018 / Published online: 16 March 2018
© The Author(s) under exclusive licence to The Japan Society of Human Genetics 2018

The human colonization and adaptation of Tibetan Plateau attract wide concerns in recent years. To date, the permanent human occupation of Tibetan Plateau is a topic of hot dispute in archaeology [1, 2]. Massive genetic studies have been conducted to address this issue by estimating the divergence time between highland Tibetans and lowland populations (e.g., Han; Table S1). The comprehensive analyses of mitochondrial DNA and Y chromosome markers revealed at least two distinct migratory waves into the Tibetan Plateau. The early Upper Paleolithic migration was dated ~30 kilo years ago (kya); in addition, the later Neolithic colonization was traced back to 10–7 kya [3, 4]. By analyzing data of exome and genome-wide single-nucleotide polymorphisms, Tibetan and Han populations were proposed to diverge 2750 [5] and 4725 years ago [6], respectively. The two recent studies based on whole-genome re-sequencing dated this divergence to 15–9 kya [7] and 58–44 kya [8], respectively. Thus, the time for

divergence between Tibetan and other populations is still in chaos.

To address the issue, we employ the multiple sequential Markovian coalescent (MSMC) [9] approach to infer the demographic history of Tibetan and other lowland populations in East Asia. The high-depth (>30×) genomic sequencing data (Table S2) include Tibetan ($n = 8$, newly generated in this study), Han ($n = 3$), Tu ($n = 2$), Mongolian ($n = 2$), and Japanese ($n = 3$) [10]. We perform calculation for eight (i.e., four diploid genomes; Fig. 1c) and four (i.e., two diploid genomes; Fig. 1b) haplotypes, respectively. These computations are repeated with another independent sampling and the results are largely consistent (Figure S1,S2). We find that MSMC based on eight haplotypes presents more coalescent events than that based on four haplotypes. When considering four haplotypes, our MSMC results show two kinds of splits: the early one is 23–15 kya indicated by Han-Japanese and Tibetan-Han; the later one is 19–11 kya represented by Tibetan-Japanese, Tibetan-Mongolian, and Tibetan-Tu (Fig. 1b). It is unable to further discern the divergence differences within the two major splits (e.g., Han-Japanese and Tibetan-Han). As compared the MSMC results of four haplotypes, each pair of population divergence can be identified based on eight haplotypes (Fig. 1c), suggesting that genomic data of eight haplotypes are more informative to increase the molecular resolution. Confusingly, the inferred divergence patterns of eight haplotypes were different from those of four haplotypes. As for the Tibetan population, the early split inferred by using eight and four haplotypes is from Japanese at 25–20 kya and from Han at 23–15 kya, respectively. Both calculations suggest the Tibetan having closest relationship with the Mongolian population. The Tibetan-Mongolian divergence is dated as 19–11 and 16–10 kya by using eight and four haplotypes, respectively.

When did Tibetan split from other populations? We cannot give a clear answer to this question based on current analyses of genomic data. Indeed, our results can provide some caveats here. Because of the relative high computation

These authors contributed equally: Xing-Yan Yang, and Shan-Shan Dai.

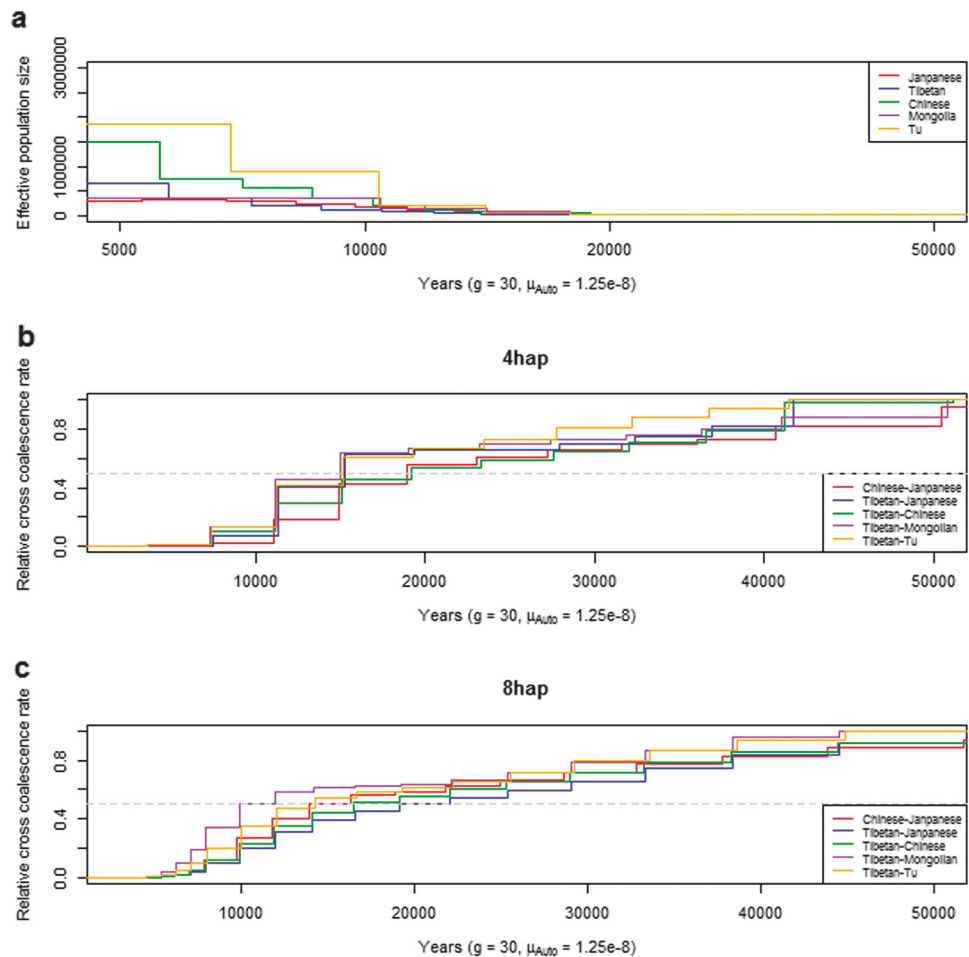
Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s10038-018-0424-0>) contains supplementary material, which is available to authorized users.

✉ Min-Sheng Peng
pengminsheng@mail.kiz.ac.cn

✉ Ya-Ping Zhang
zhangyp@mail.kiz.ac.cn

- 1 State Key Laboratory for Conservation and Utilization of Bioresources in Yunnan, Yunnan University, Kunming, China
- 2 State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
- 3 Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China
- 4 KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming, China

Fig. 1 The historical effective population size and divergence time of Panel 1. **a** The historical effective population size. **b** Divergence time based on Panel 1 with four haplotypes. **c** Divergence time based on Panel 1 with eight haplotypes. Four haplotypes refer to a pair of one highlander and one lowlander; eight haplotypes refers to a pair of two highlanders and two lowlanders. An autosomal neutral mutation rate $\mu_{\text{Auto}} = 1.25 \times 10^{-8}$ per base per human generation for a generation time of 30 years was used. See Table S2 for the information of Panel 1



load [9], MSMC is usually run based on two or four haplotypes [7, 8, 10]. Our results suggest that sample size plays substantial roles in MSMC analyses. How to interpret the differences in the inference of population relationship and divergence time based on eight and four haplotypes is still unclear. Using eight haplotypes, which contain more genetic information, could be a preferential choice. In addition, MSMC cross coalescence rates behave roughly linear under admixture [11]. The investigation of genetic admixture in individuals used in MSMC analysis may be required. Taken above, the uncertainty of population relationship as well as divergence time between Tibetan and other populations should be considered, especially in demographic inference.

Availability of data and materials

The data reported in this study are available in the Genome Sequence Archive (GSA) in Big Data Center, Beijing Institute of Genomics (BIG), and Chinese Academy of Science (<http://gsa.big.ac.cn>; accession numbers PRJCA000600).

Acknowledgements This study was supported by the Bureau of Science and Technology of Yunnan Province and the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (the Large Research Infrastructure Funding).

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Zhang DD, Li S-H. Comment on “Permanent human occupation of the central Tibetan Plateau in the early Holocene”. *Science*. 2017;355:64–7.
- Jade d’Alpoim Guedes RKyleBocinsky, Butler EE. Comment on “Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P. *Science*. 2015;348:872.
- Shi H, Zhong H, Peng Y, Dong YL, Qi XB, Zhang F, et al. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol*. 2008;6:45.
- Qi X, Cui C, Peng Y, Zhang X, Yang Z, Zhong H, et al. Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the tibetan plateau. *Mol Biol Evol*. 2013;30:1761–78.

5. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*. 2010;329:75–8.
6. Yang J, Jin ZB, Chen J, Huang XF, Li XM, Liang YB, et al. Genetic signatures of high-altitude adaptation in Tibetans. *Proc Natl Acad Sci. USA*. 2017;114:4189–94.
7. Lu D, Lou H, Yuan K, Wang X, Wang Y, Zhang C, et al. Ancestral origins and genetic history of tibetan highlanders. *Am J Hum Genet*. 2016;99:580–94.
8. Hu H, Petousi N, Glusman G, Yu Y, Bohlender R, Tashi T, et al. Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLoS Genet*. 2017;13:e1006675.
9. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 2014;46:919–25.
10. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*. 2016;538:201–6.
11. Pagani L, Lawson DJ, Jagoda E, Morseburg A, Eriksson A, Mitt M, et al. Genomic analyses inform on migration events during the peopling of Eurasia. *Nature*. 2016;538:238–42.