



EDITORIAL

T2T-YAO, T2T-SHUN, and more

Jingfa Xiao^{1,2,3}, Jun Yu^{2,3,*}

¹ National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China

² CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

Received 4 September 2023; accepted 18 September 2023

Available online 22 September 2023

The first high-quality reference genome for Chinese populations has just been released [1], T2Y-YAO (尧yáo), and it starts a first-of-its-kind series for building virtual ancestor genomes (VAGs) for Chinese population-based studies and healthcare applications [2,3]. The reasons are multifold. First, the Han people (汉人 hàn rén) are the world's largest ethnicity, ~ 1.5 billion in total, ~ 19% of the global population, and 91% of the Chinese nationals. Second, the Chinese population history — regardless of whether the precise time period of each nation and State has or has not been determined — and complexity are being recorded faithfully in their genome sequences. Third, the Han ethnic groups are still co-habiting with other 55 minor ethnic peoples without geopolitical and geographical barriers.

In these obvious regards, we have yet to construct a reference genome-based network in which the edges and nodes are VAGs so connected based on between- and within-population sequence variations that are supported by largely within-population data (Figure 1). In other words, we need to choose historically representative geographical sites first and to assemble as many “nodes” — historically- or geographically-defined edge or reference genomes (HEGs) — as possible, and then to connect the neighboring HEGs to build a network. As the network and graphic elements grow in size and number, respectively, HEGs will be consolidated into VAGs. In the end, these VAGs will be tailored to the history of population migration and mixing, as well as their relevant timing, and most importantly, they can be used for effective genome-wide asso-

ciation study (GWAS) analyses for phenotype and disease/symptom mapping when the size of the inheritance-limited population reaches an expected significance. It is anticipated that thousands of HEGs and hundreds of VAGs are to be acquired for mapping useful variations that are disease- or health-relevant [4]. Therefore, we expect an effort to build such a network to be firmly supported within a few years [5], which serves as a database for population-centric studies as genome sequencing becomes affordable not only in rich counties but also in developing counties together with an awareness of the value of precision medicine to all human societies.

The purpose of expanding VAG-based network is to map genome variations effectively and precisely. One can imagine that an unknown genome sequence is better mapped to that of its close relatives rather than a distant one in which noises from between-population variations become substantially interruptive. Such a mapping process is often time-consuming, so it needs constant care for effectiveness and efficiency. Furthermore, within the next ten years or so, thousands and millions of genomes from most individuals will be acquired and the job of collecting, processing, and utilizing these data will definitely create enormous bottlenecks, let alone interpreting and explaining them to the doctors and the general public.

While we are welcoming the first HEG [we have yet to know if it will become one of the historically-defined ancestral genomes (HAGs)], it is essential to point out the fact that not only T2T-SHUN (舜shùn) and T2T-YU (禹yǔ), where SHUN and YU are well-known Chinese ancestors different from

* Corresponding author.

E-mail: junyu@big.ac.cn (Yu J).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2023.09.002>

1672-0229 © 2023 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

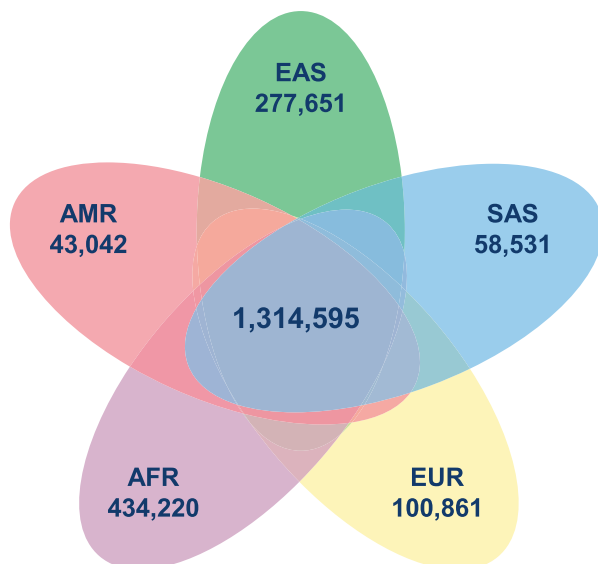
No. of shared major alleles (allele frequency ≥ 0.5)

Figure 1 Distribution of major alleles among five major ethnic groups

Data were extracted from the 1000 Genomes Project. The number indicates the count of major alleles (allele frequency ≥ 0.5) shared within each of the five major ethnic groups or among all of them. EAS, East Asian; AMR, admixed American; AFR, African; EUR, European; SAS, South Asian.

YAO, are needed, but also several HEGs from each of the 55 Chinese ethnicities are to be expected!

Competing interests

Both authors declare no competing interests.

CRedit authorship contribution statement

Jingfa Xiao: Data curation, Writing – review & editing. **Jun Yu:** Conceptualization, Data curation, Writing – original draft, Writing – review & editing, Supervision. Both authors have read and approved the final manuscript.

ORCID

ORCID 0000-0002-2835-4340 (Jingfa Xiao)

ORCID 0000-0002-2702-055X (Jun Yu)

References

- [1] He Y, Chu Y, Guo S, Hu J, Li R, Zheng Y, et al. T2T-YAO: a telomere-to-telomere assembled diploid reference genome for Han Chinese. *Genomics Proteomics Bioinformatics* 2023;21:1085–100.
- [2] Du Z, Ma L, Qu H, Chen W, Zhang B, Lu X, et al. Whole genome analyses of Chinese population and *de novo* assembly of a Northern Han genome. *Genomics Proteomics Bioinformatics* 2019;17:229–47.
- [3] Ling Y, Jin Z, Su M, Zhong J, Zhao Y, Yu J, et al. VCGDB: a dynamic genome database of the Chinese population. *BMC Genomics* 2014;15:265.
- [4] Yu J, Hu S. On the ultimate finishing line of the Human Genome Project. *Innovation (Camb)* 2021;2:100133.
- [5] Gao Y, Yang X, Chen H, Tan X, Yang Z, Deng L, et al. A pangenome reference of 36 Chinese populations. *Nature* 2023;619:112–21.