

Differentiated adaptative genetic architecture and language-related demographical history in South China inferred from 619 genomes from 56 populations

Qiuxia Sun^{1,2,*}, Mengge Wang^{2,14,*#}, Tao Lu^{1,*}, Shuhan Duan^{3,2}, Yan Liu^{3,2}, Jing Chen^{2,5}, Zhiyong Wang^{2,6}, Yuntao Sun^{2,7}, Xiangping Li^{2,6}, Shaomei Wang^{2,8}, Liuyi Lu^{4,2}, Liping Hu⁶, Libing Yun⁷, Junbao Yang⁴, Jiangwei Yan⁵, Shengjie Nie⁶, Yanfeng Zhu⁸, Gang Chen¹², Chuan-Chao Wang¹³, Chao Liu^{9,10,11}, Guanglin He^{2,14,*#}, Renkuan Tang^{1,#}

Correspondence: Mengge Wang (Menggewang2021@163.com), Renkuan Tang (renktang2012@163.com), Guanglin He (guanglinhescu@163.com)

*Qiuxia Sun, Mengge Wang, Tao Lu and Guanglin He contributed equally to this work

¹Department of Forensic Medicine, College of Basic Medicine, Chongqing Medical University, Chongqing, 400331, China

²Institute of Rare Diseases, West China Hospital of Sichuan University, Sichuan University, Chengdu, 610000, China

³School of Basic Medical Sciences, North Sichuan Medical College, Nanchong, 637100, China

⁴School of Clinical Medical Sciences, North Sichuan Medical College, Nanchong, 637100, China

⁵School of Forensic Medicine, Shanxi Medical University, Jinzhong, 030001, China

⁶School of Forensic Medicine, Kunming Medical University, Kunming, 650500, China

⁷West China School of Basic Science & Forensic Medicine, Sichuan University, Chengdu, 610041, China

⁸Department of Public Health, Chengdu Medical College, Chengdu, 610500, China

⁹Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, 510275, China

¹⁰Guangzhou Forensic Science Institute, Guangzhou, 510055, China

¹¹Anti-Drug Technology Center of Guangdong Province, Guangzhou, 510230, China

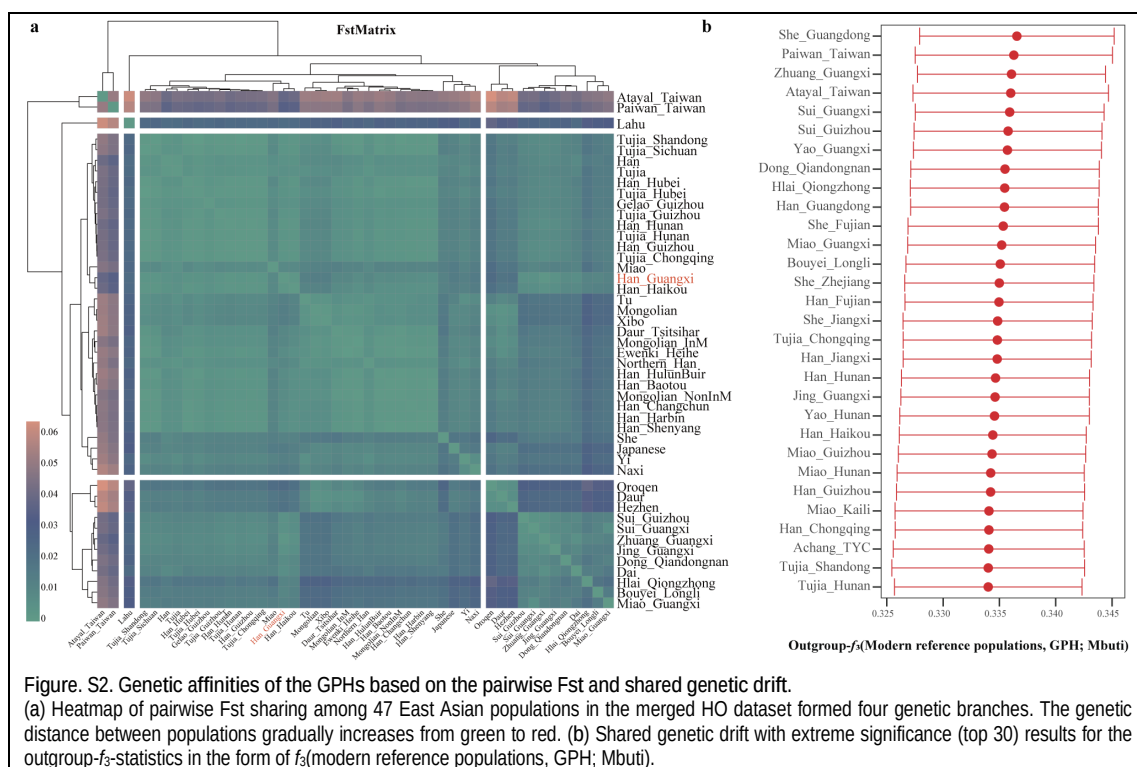
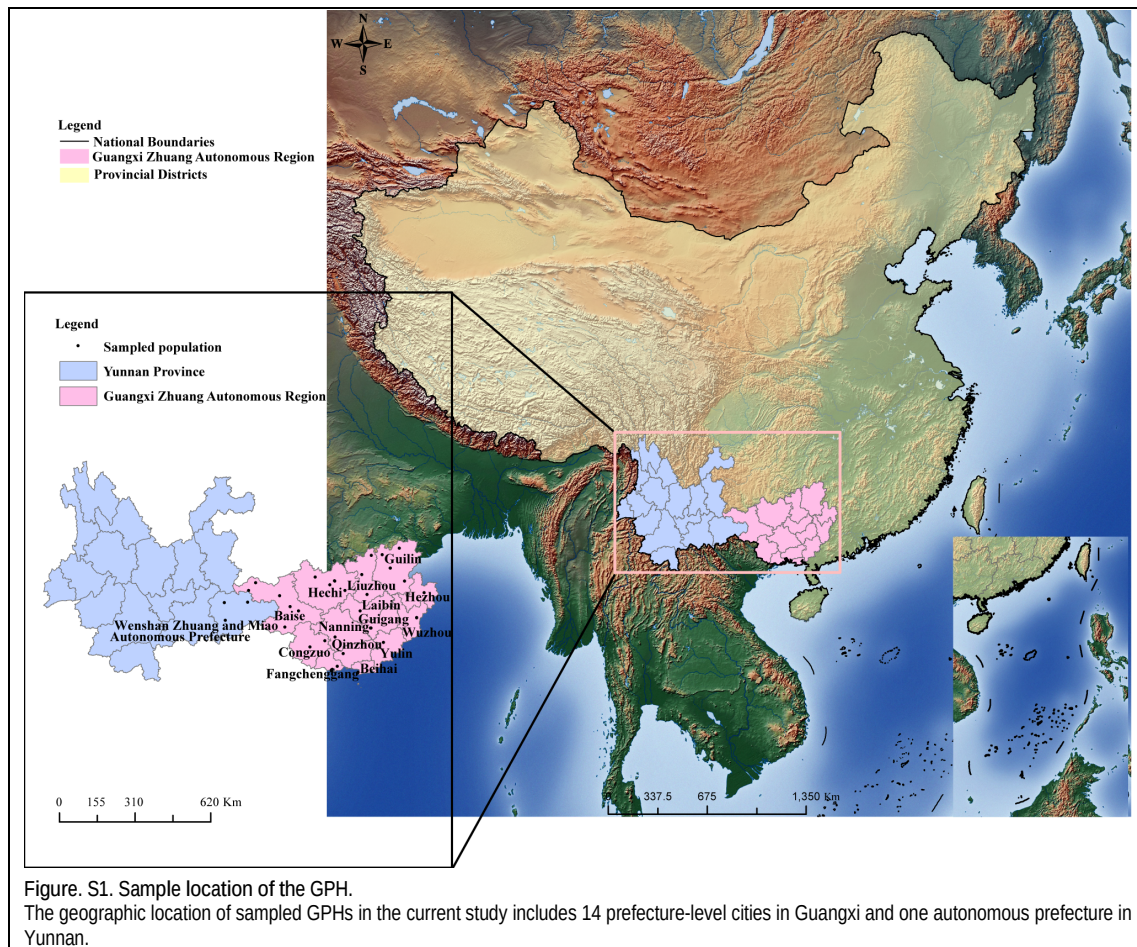
¹²Hunan Key Lab of Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410075, China

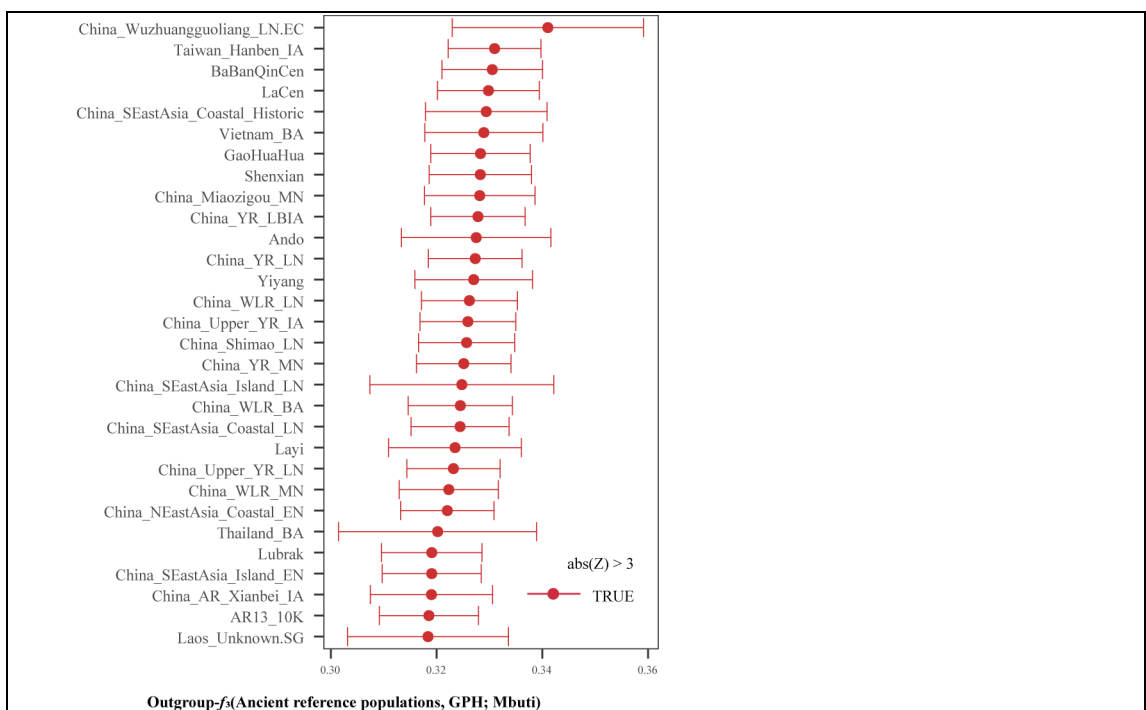
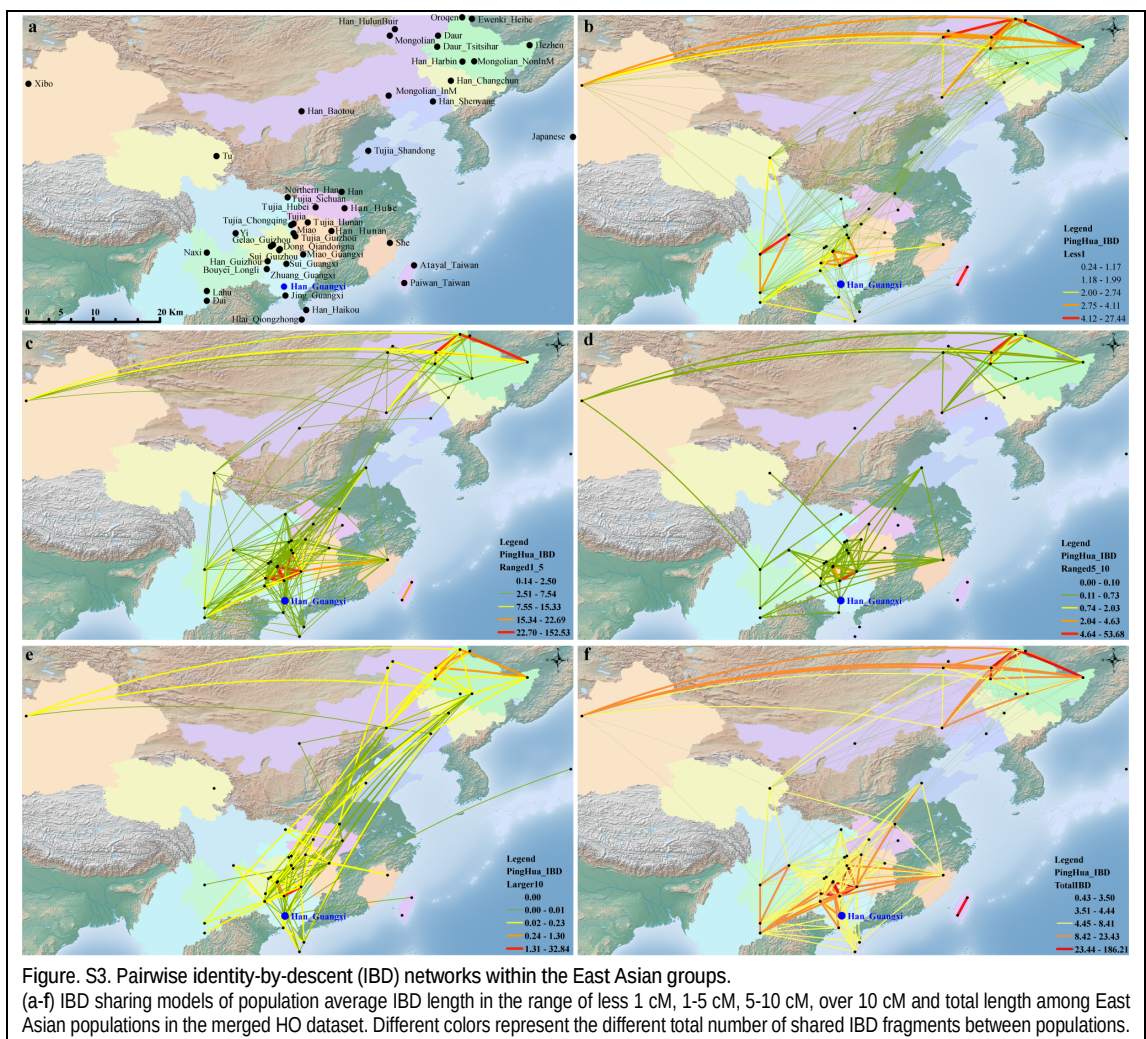
¹³State Key Laboratory of Cellular Stress Biology, National Institute for Data Science in Health and Medicine, School of Life Sciences, Xiamen University, Xiamen 361005, Fujian, China

¹⁴Center for Archaeological Science, Sichuan University, Chengdu, 610000, China

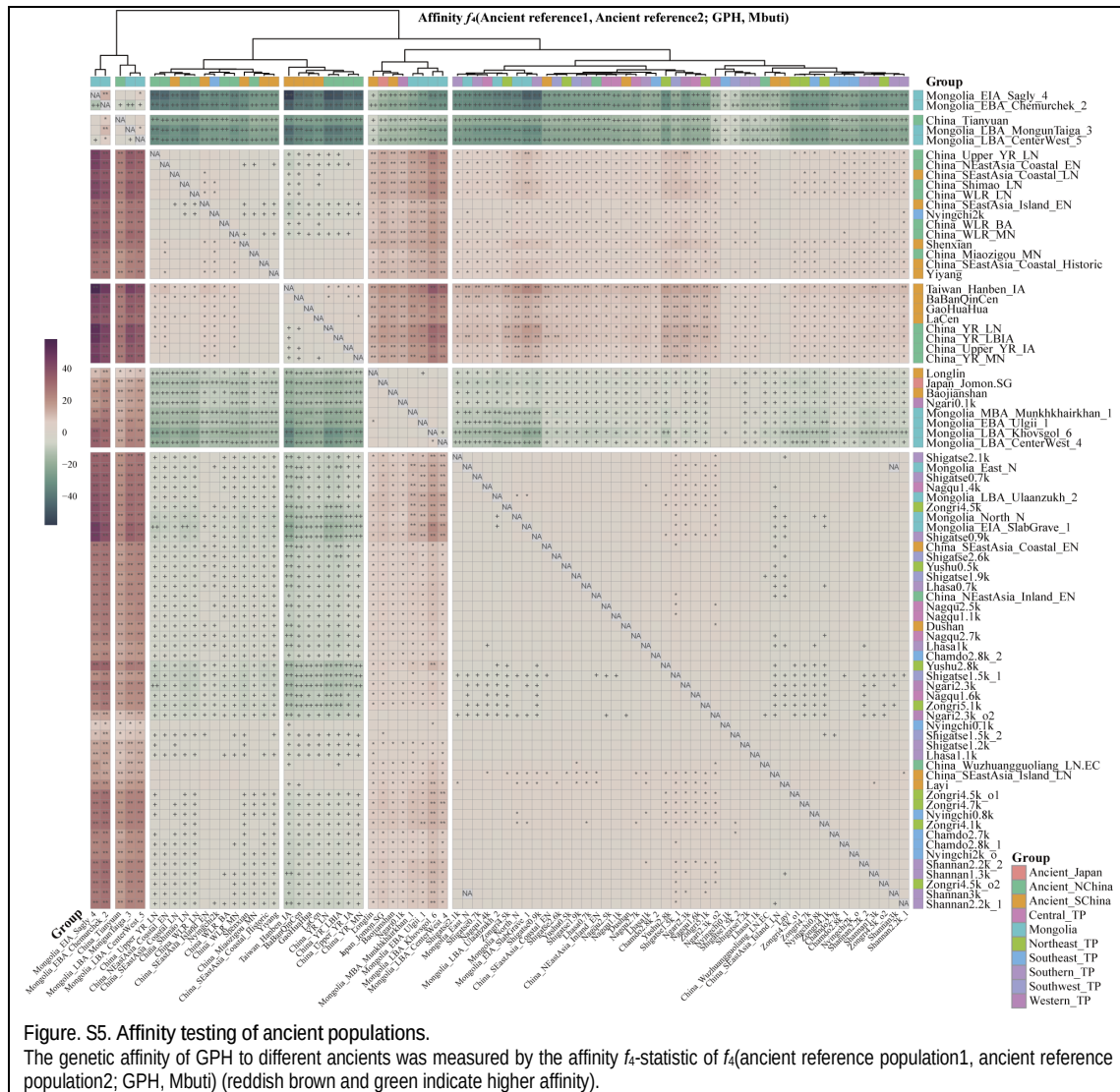
Contents

Figure. S1. Sample location of the GPH.	3
Figure. S2. Genetic affinities of the GPHs based on the pairwise F_{st} and shared genetic drift.	3
Figure. S3. Pairwise identity-by-descent (IBD) networks within the East Asian groups.	4
Figure. S4. Genetic affinities between GPH and 30 ancient Chinese populations.	4
Figure. S5. Affinity testing of ancient populations.	5
Figure. S6. F_4 -statistic test for gene introgression in GPH.	6
Figure. S7. Asymmetric testing of GPHs.	7
Figure. S8. Allele frequency-based asymmetrical f_4 -statistics test.	7
Figure. S9. Runs of homozygosity (ROH) of the populations in East Asia.	8
Figure. S10. Inferred phylogenetic relationship using TreeMix.	9
Figure. S11. Overview of inferred admixture for GPH using fastGLOBETROTTER.	9
Figure. S12. Effective population sizes of Guangxi groups inferred from IBDNe.	10
Figure. S13. Fine-scale genetic structure and coancestry curves based on the shared haplotype data.	11
Figure. S14. The population structure of five ethnic communities and GPH in Guangxi Province.	12
Figure. S15. Paternal lineages in the context of East Asia.	12
Figure. S16. Y chromosome haplogroup distribution.	13
Figure. S17. Mitochondrial haplogroup distribution.	14
Figure. S18. Natural selection signals and highly differentiated loci.	14





Mbuti).



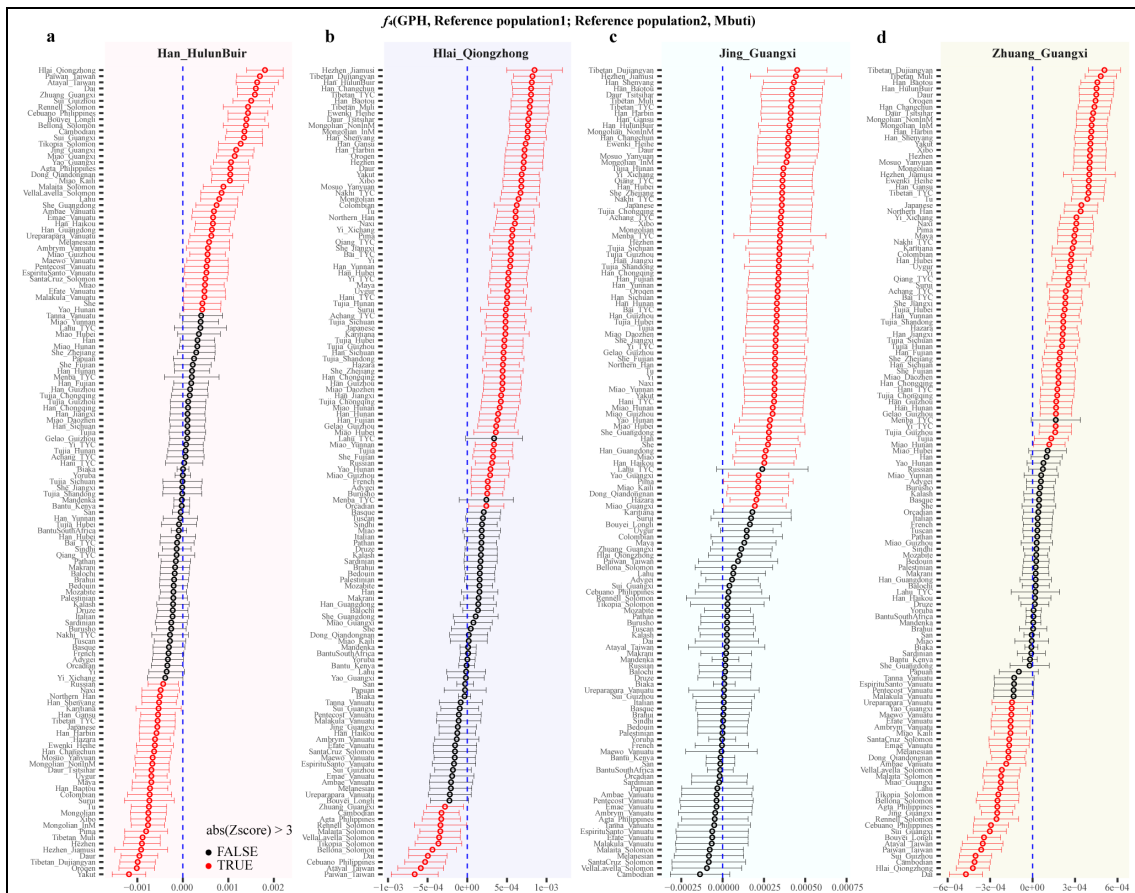


Figure. S6. F_4 -statistic test for gene introgression in GPH.

(a-d) F_4 -statistics in the form of $f_4(\text{GPH, reference population1; reference population2, Mbuti})$, where reference population1 represents the selected Northern Han (Han_HulunBuir) and Southern ethnic minorities (Hlai_Qiongzong/Zhuang/Jing_Guangxi), reference population2 represents global groups in the merged HGDP data. Different pairwise population pairs were used to explore genetic differences between GPH and reference population1.

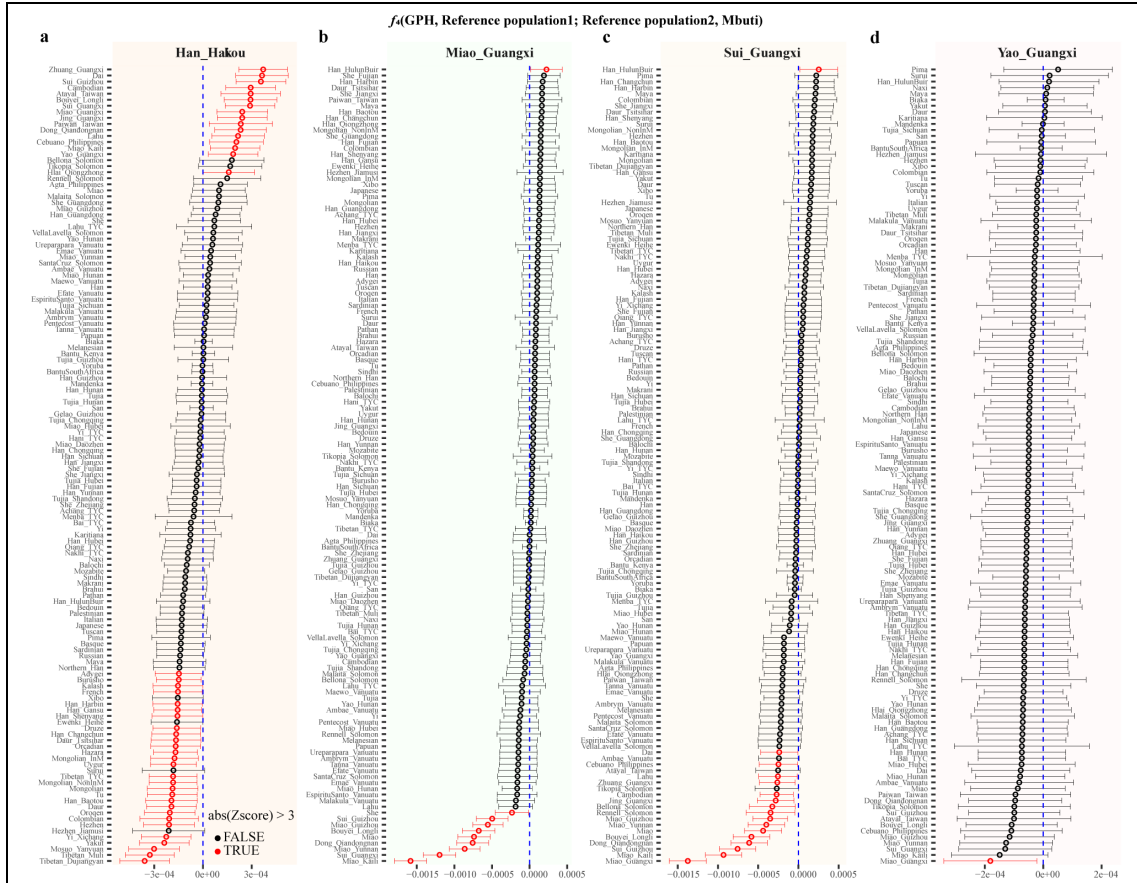


Figure. S7. Asymmetric testing of GPHs.

(a) The values of $f_4(\text{GPH, reference population1; reference population2, Mbuti})$ are shown with ± 3 standard errors, where reference population1 represents Southern Han (Han_Haikou). GPH received more alleles from Southern ethnic minorities than Han_Haikou. (b-d) The asymmetric f_4 -statistics of ethnic minority groups in Guangxi (Miao/Sui/Yao) as a target with different pairs of global reference populations. Significant negative values indicate that these groups affected more alleles from Southern ethnic minorities than those of GPH.

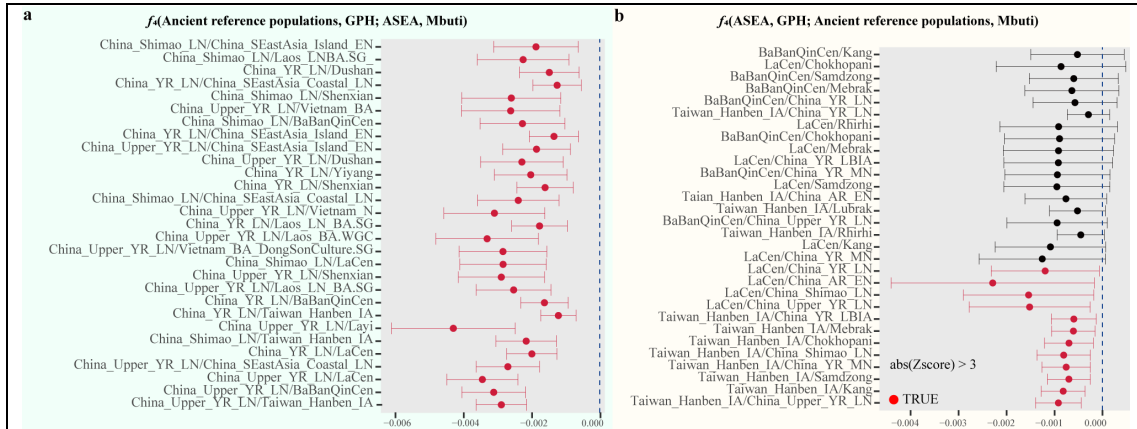
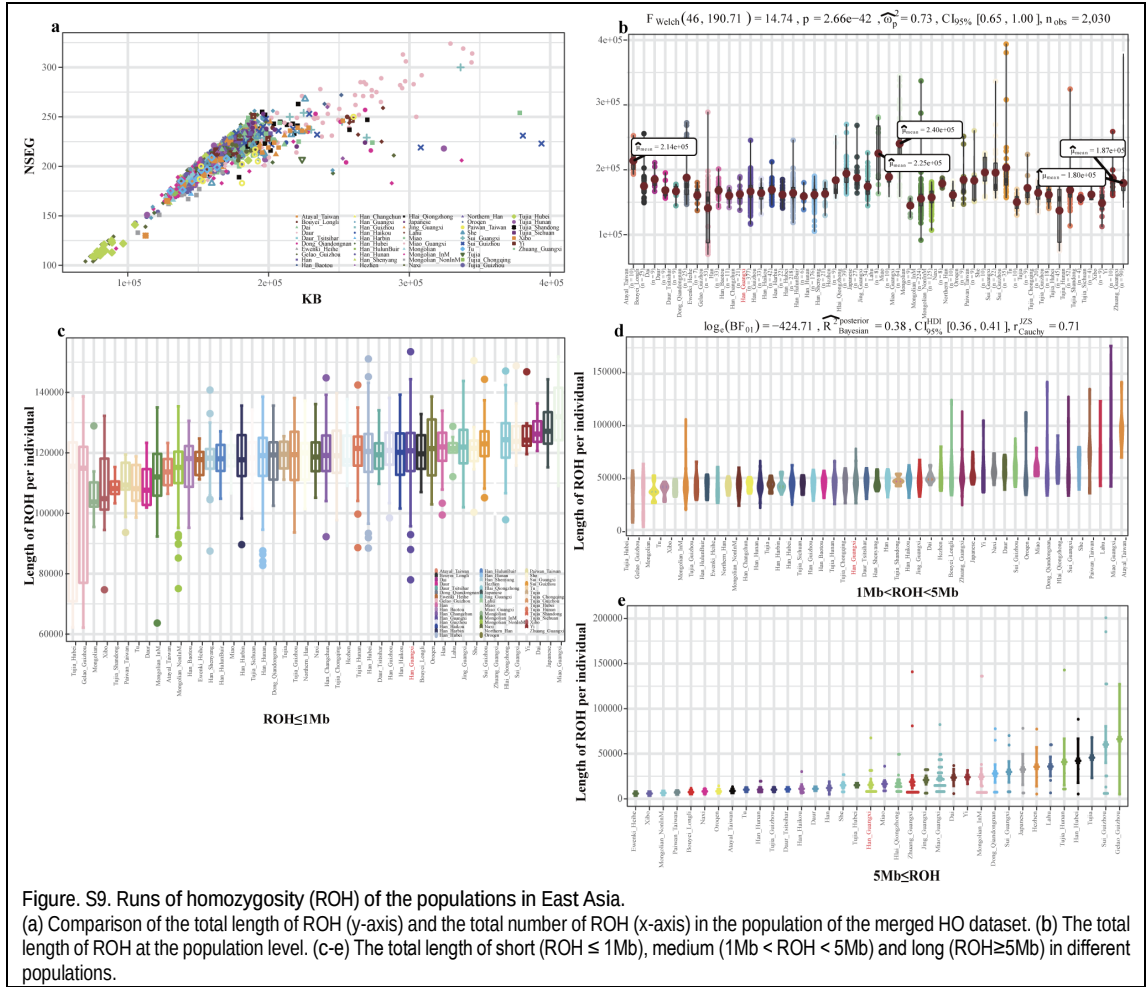
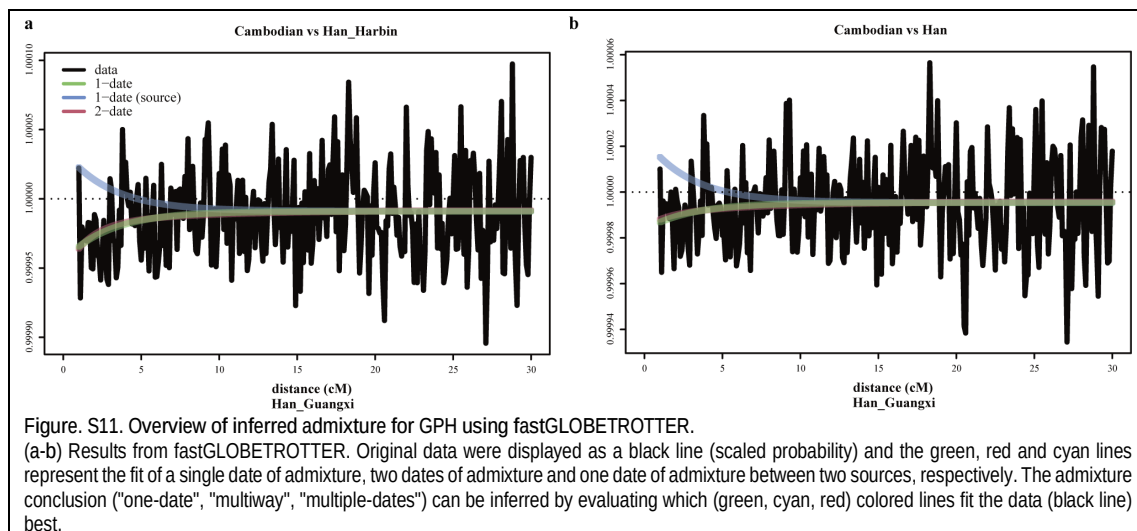
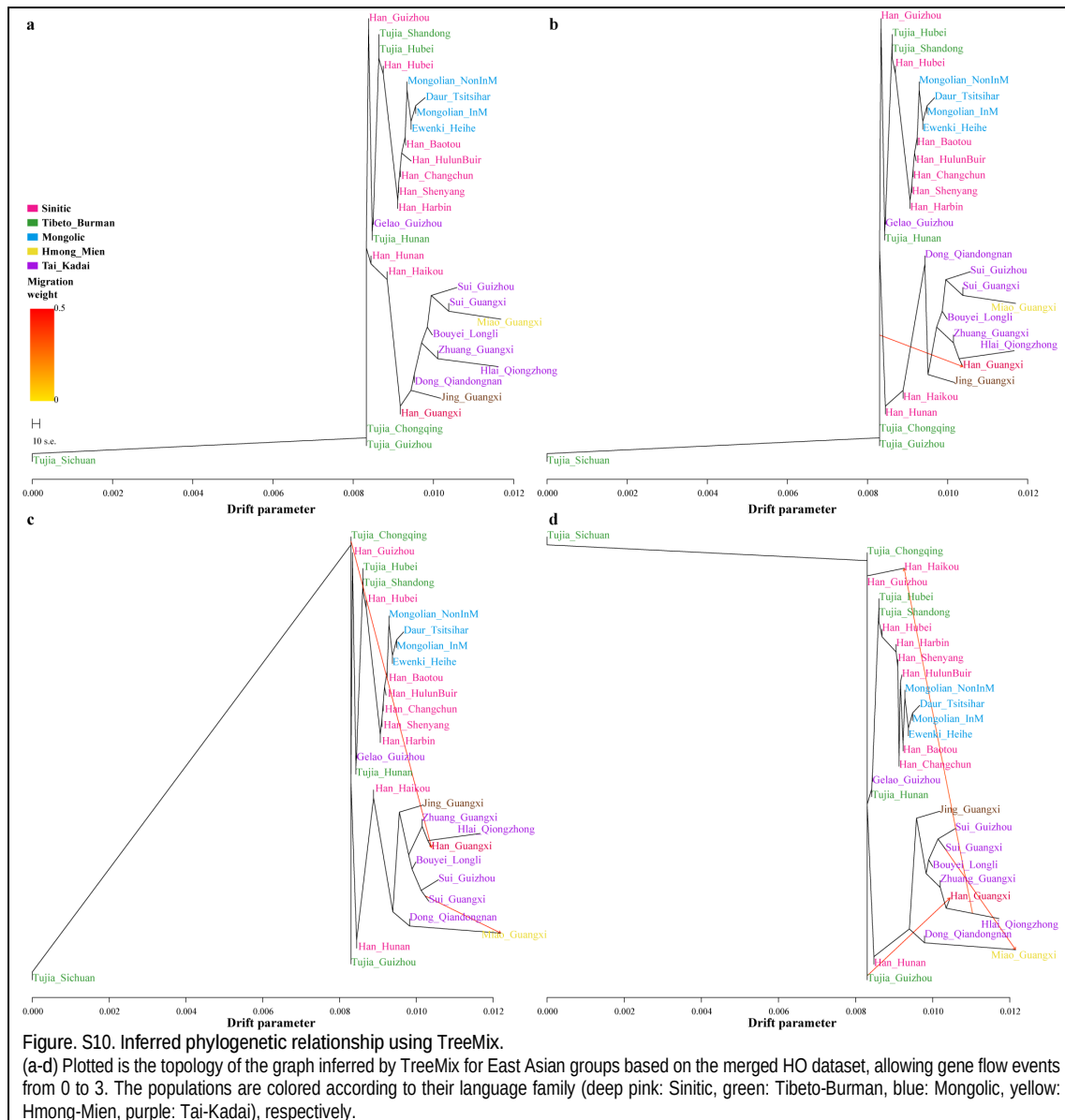
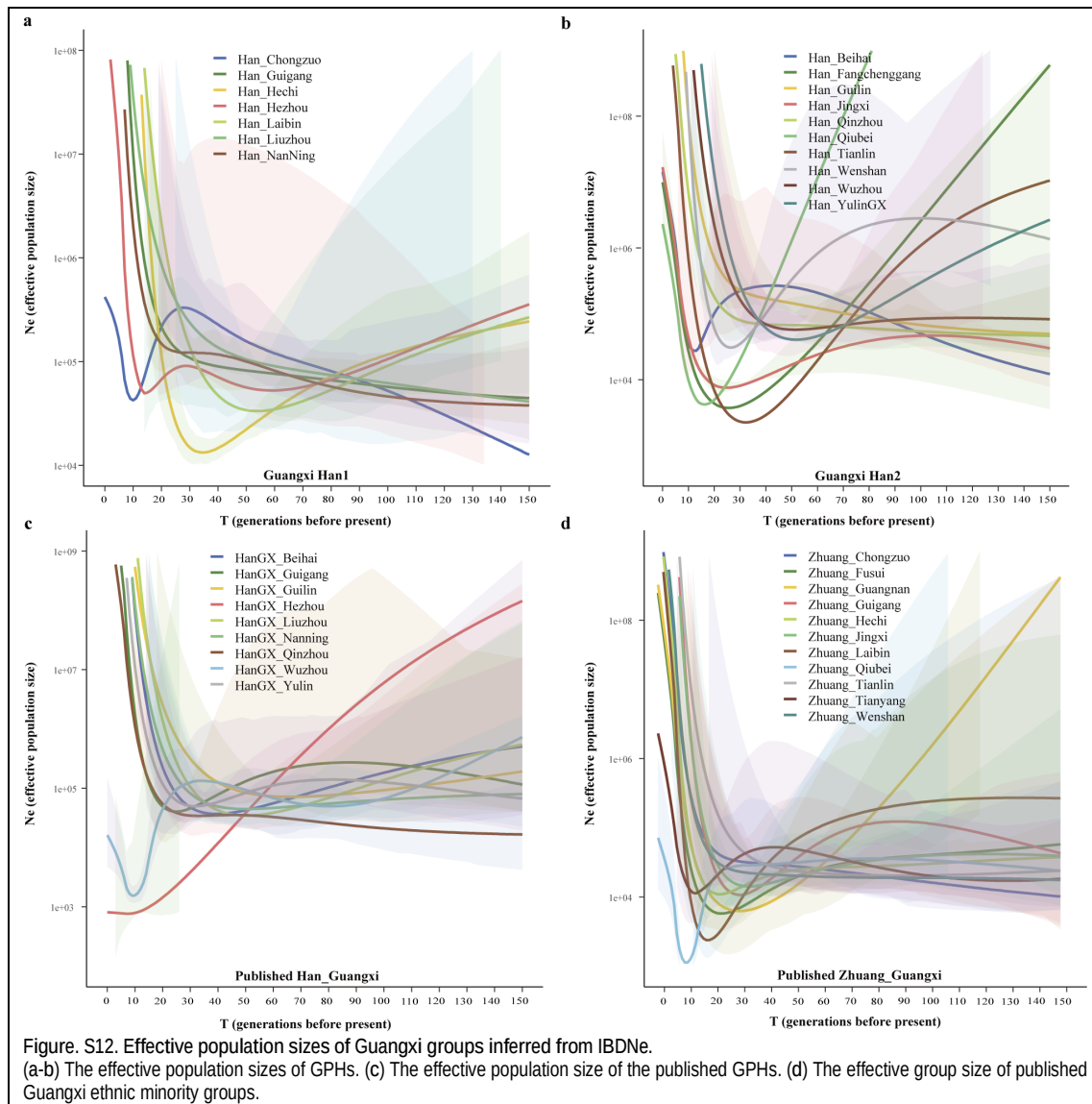


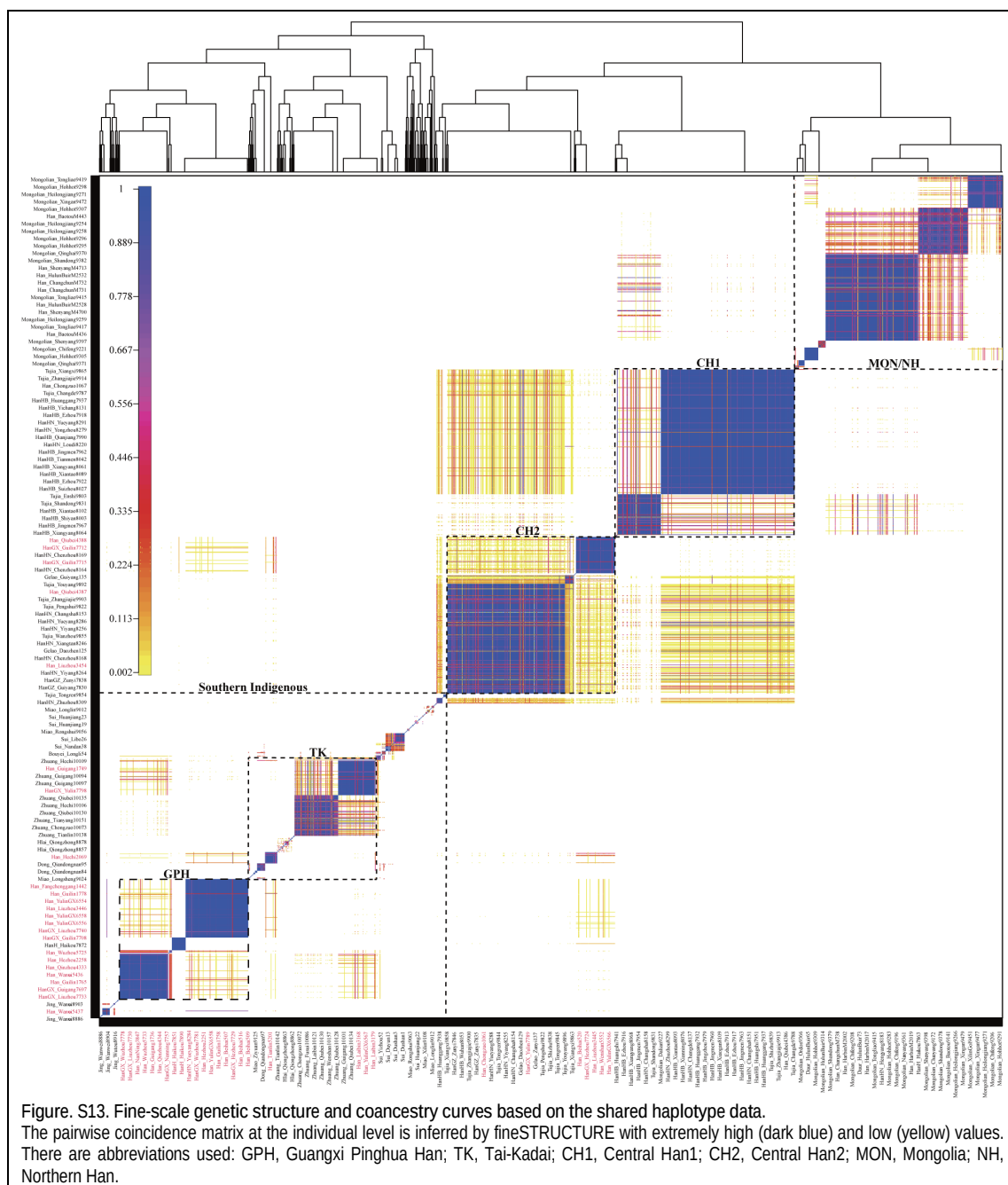
Figure. S8. Allele frequency-based asymmetrical f_4 -statistics test.

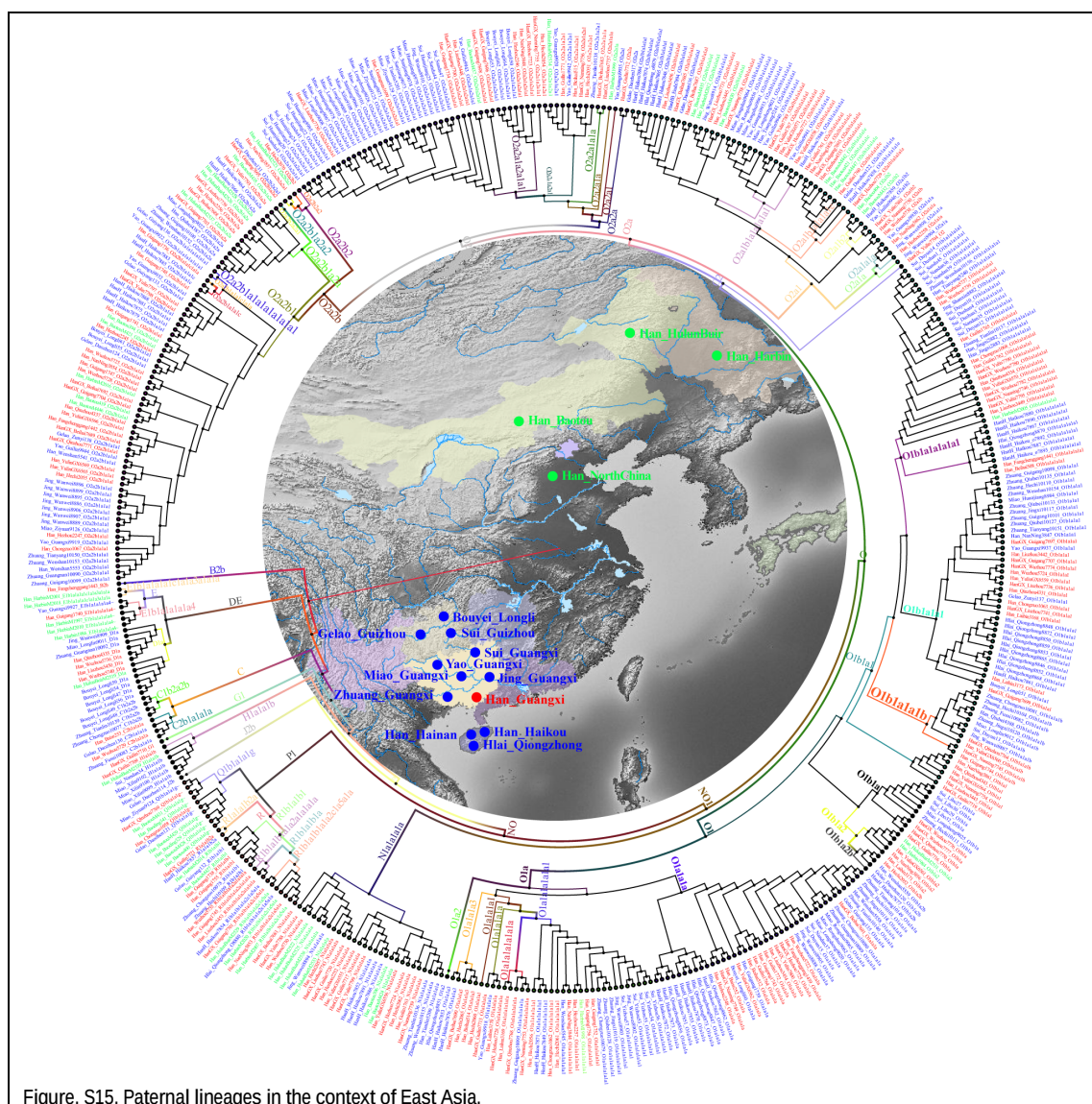
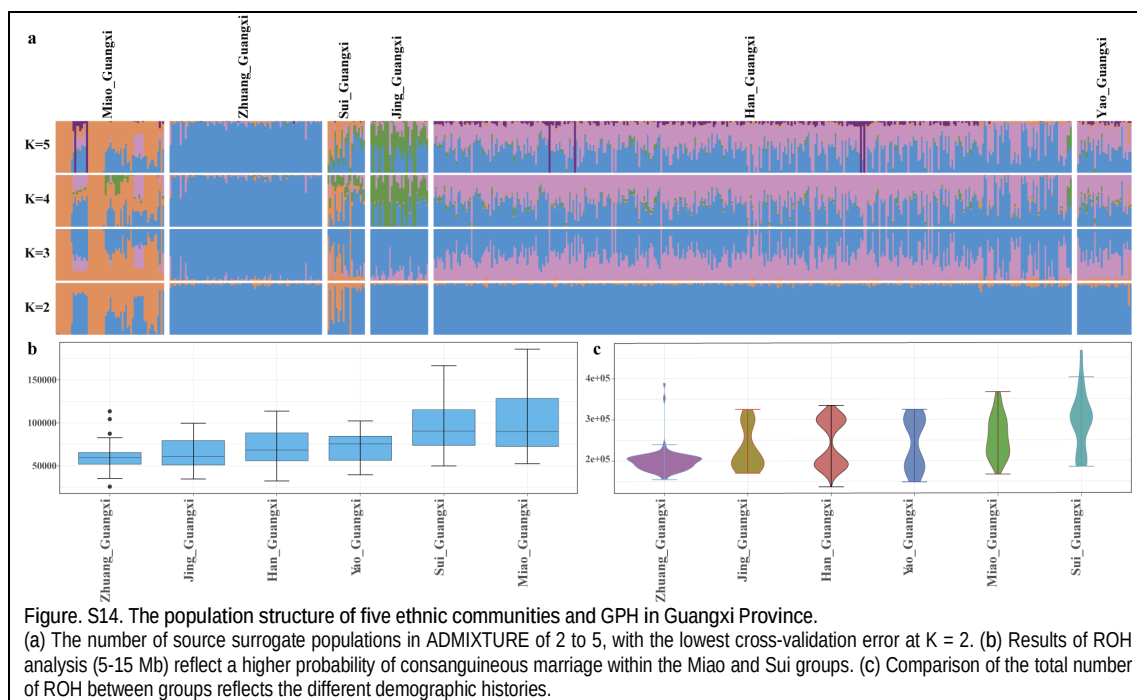
(a) F_4 -statistics in the form of $f_4(\text{ancient reference populations, GPH; ASE, Mbuti})$ testing the GPH's genetic association with ancient Southern East Asian ancestries (ASEA) (including Taiwan_Hanben_IA, BaBanQinCen, LaCen, China_SEEastAsia_Coastal_LN, Layi, Laos_LN_BA.SG, Shenxian, Vietnam_BA_DongSonCulture.SG, Laos_BA.WGC, Vietnam_N, Yiyang, Dushan, China_SEEastAsia_Island_EN and Vietnam_BA) in contrast to other ancient East Asians. (b) Numbers of GPH ancestral sources were determined by f_4 -statistics in the form of $f_4(\text{ASEA, GPH; ancient reference populations, Mbuti})$, where ASEA represents an identified ancestor (Taiwan_Hanben_IA, LaCen and BaBanQinCen) of GPH, ancient reference populations represent other ancient populations (the ancient people of East Asia, Southeast Asia and South Asia).











ML tree of 474 chrY sequences. Colors in the different groups correspond to the geographic distribution of the samples (green: Northern Chinese groups, dark blue: Southern Chinese groups, red: GPH). Different haplotypes were colored with different colors randomly.

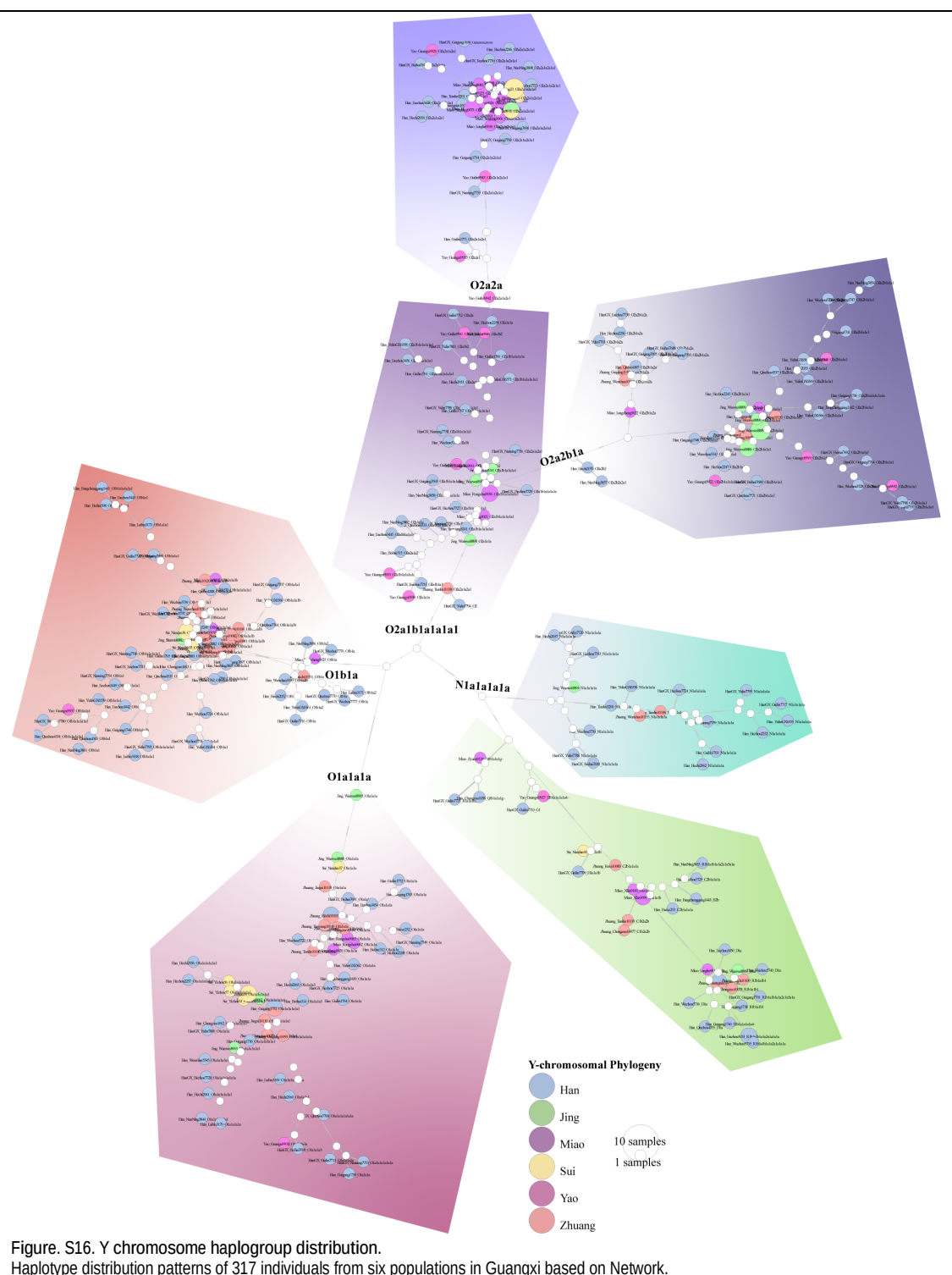


Figure. S16. Y chromosome haplogroup distribution. Haplotype distribution patterns of 317 individuals from six populations in Guangxi based on Network.

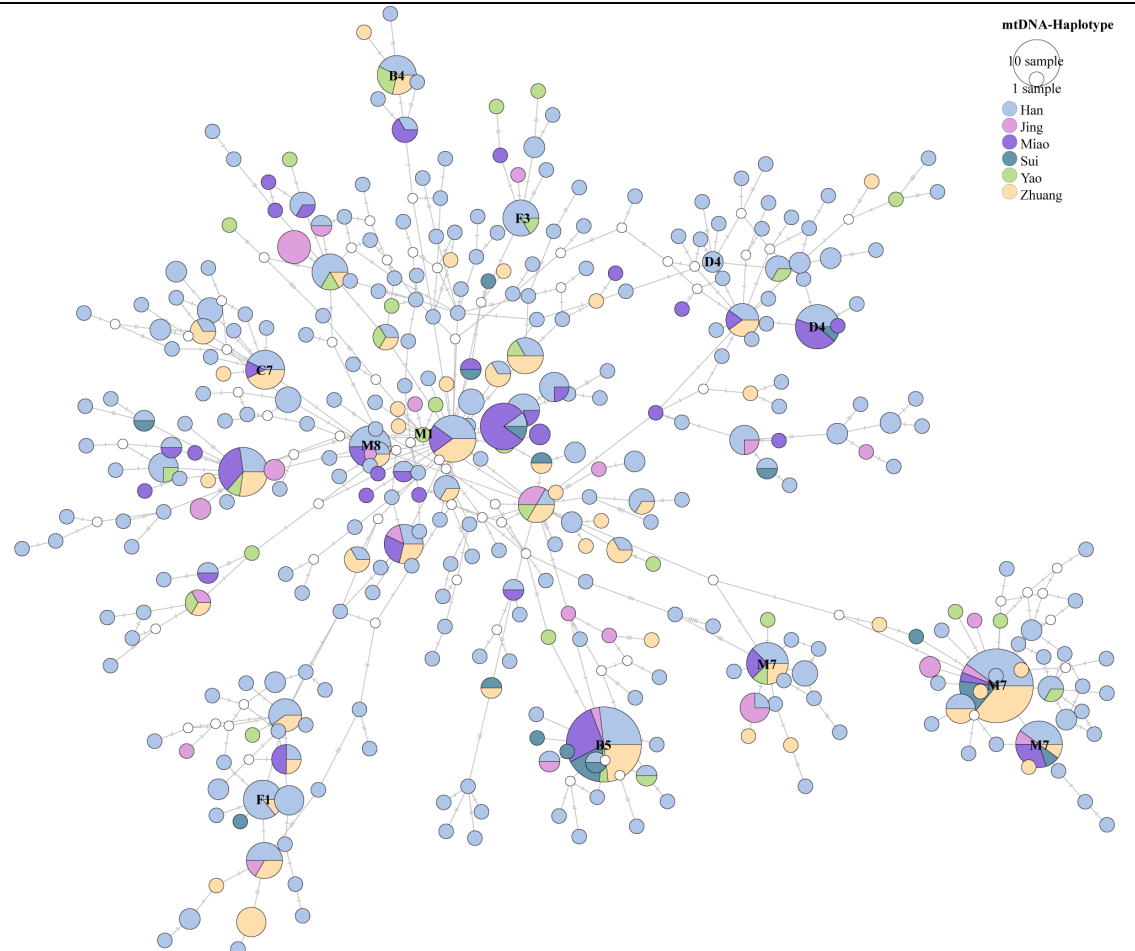


Figure. S17.Mitochondrial haplogroup distribution.
The maternal genetic background of 619 individuals from six groups in Guangxi was revealed using network-based phylogenetic relationships.

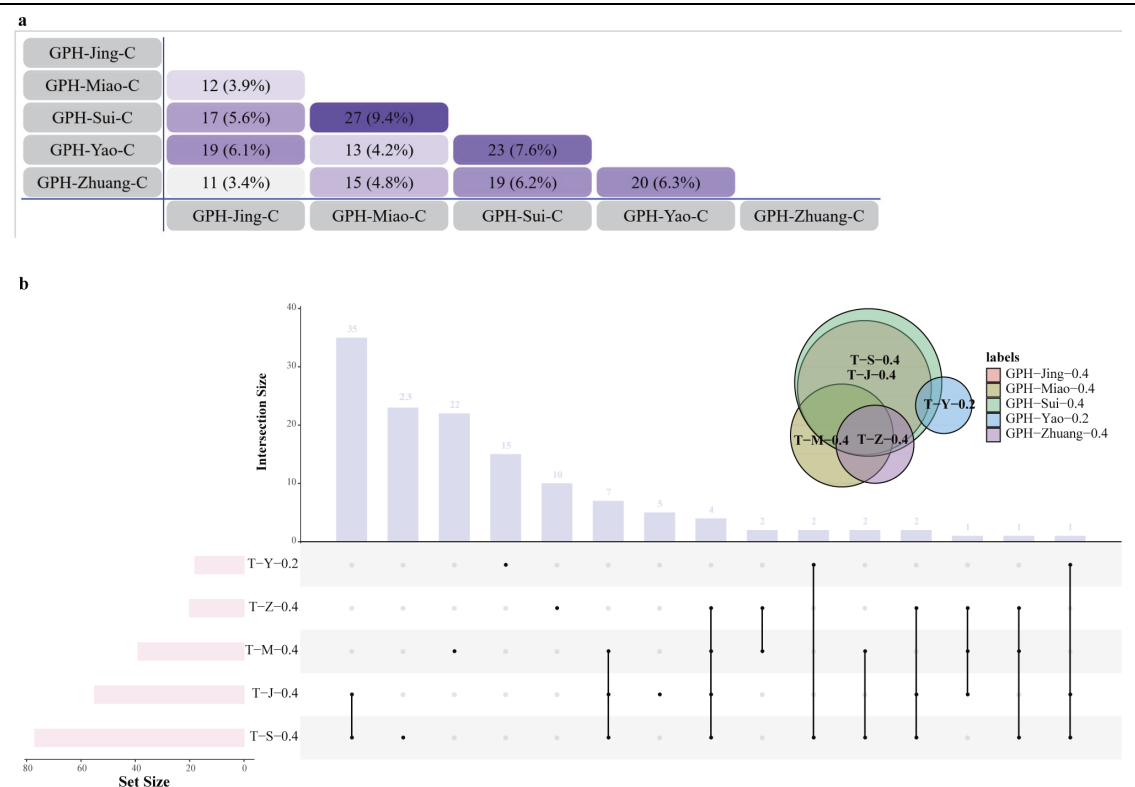


Figure. S18. Natural selection signals and highly differentiated loci.

(a) Venn diagram of regional-specific natural selection signals detected by PBS method (GPH-Jing/Miao/Sui/Yao/Zhuang_Guangxi-Han_Changchun). The number and percentage of intersections of different results are shown. See Additional file 2, Table S12 for more details.

(b) Upset plots of highly differentiated loci. Loci with differentiation frequencies higher than 0.4 between populations were selected for comparison (except for the T-Y group, which was 0.2). T, GPH; Y, Yao_Guangxi; Z, Zhuang_Guangxi; M, Miao_Guangxi; J, Jing_Guangxi; S, Sui_Guangxi; C, Han_Changchun.