

Original Research

Medieval genomes from eastern Mongolia share a stable genetic profile over a millennium

Juhyeon Lee 1, Takehiro Sato 2, Atsushi Tajima 3, Tsend Amgalantugs 4, Batmunkh Tsogtbaatar 4, Shigeki Nakagome 5,6, Toshihiko Miyake 7, Noriyuki Shiraishi 8,* , Choongwon Jeong 1,* , Takashi Gakuhari 6,*

1 School of Biological Sciences, Seoul National University, Seoul, 08826, Republic of Korea

2 Graduate School of Medicine, University of the Ryukyus, Okinawa, 903-0213, Japan

3 Faculty of Medicine, Kanazawa University, Kanazawa, 920-0934, Japan

4 Institute of Archaeology, Mongolian Academy of Sciences, Ulaanbaatar, 13330, Mongolia

5 School of Medicine, Trinity College Dublin, Dublin, 2, Ireland

6 Institute for the study of Ancient Civilization and Cultural Resources, College of Human and Social Sciences, Kanazawa University, Kanazawa, 920-1192, Japan

7 School of History, College of Humanities, Shukutoku University, Tokyo, 174-8631, Japan

8 Faculty of Humanities, Niigata University, Niigata, 950-2181, Japan

* **Correspondence:** Noriyuki Shiraishi; Email: shira@human.niigata-u.ac.jp; Choongwon Jeong; Email: cwjeong@snu.ac.kr; Takashi Gakuhari; Email: gakuhari@staff.kanazawa-u.ac.jp

Supplementary Materials

Contents

1. Figure S1. The post-mortem chemical damage pattern of GD and TK individuals.
2. Figure S2. The ratio of the coverage of sex chromosomes to the autosomes.
3. Figure S3. Genetic relatives detected within GD and TK sites.
4. Figure S4. Outgroup-f3 statistics of GD and TK groups.
5. Figure S5. HapROH results for GD/TK individuals.
6. Figure S6. The genetic profiles of the 6 source populations for GD/TK individuals along the two Eurasian PCs.
7. Figure S7. ADMIXTURE cross-validation error analysis for K (K=2-10) with ten replicates.
8. Figure S8. ADMIXTURE results for 1,774 present-day Eurasian and 207 ancient individuals for K=2 to K=10.
9. Figure S9. The distribution of the mean values of PC1s of four randomly sampled individuals with identical time period composition to GD.
10. Figure S10. Low heterogeneity of the genetic profiles of the four Mongol-period TK individuals.
11. Table S1. Summary of GD and TK individuals analyzed in this study.
12. Table S2. The published populations used for the population genetic analysis.
13. Table S3. Genetic relatives in this study.
14. Table S4. Summary of uniparental haplogroups of the previously published ancient individuals during Xiongnu to Mongol period.
15. Table S5. Population-based outgroup f3 results of GD and TK populations.
16. Table S6. QpAdm modeling of GD and TK individuals.

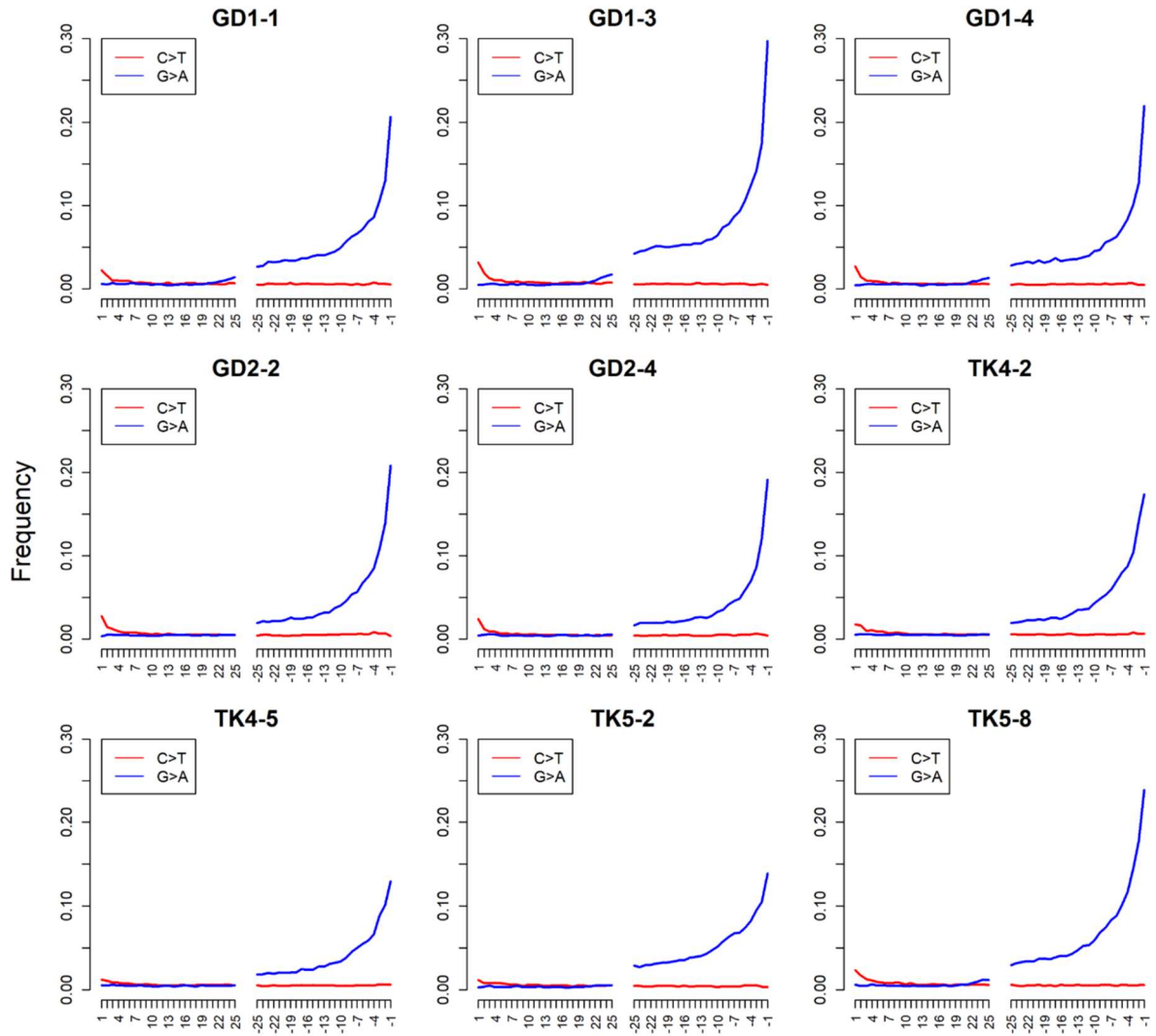


Figure S1. The post-mortem chemical damage pattern of GD and TK individuals. For each GD and TK individual, we present the base mismatch proportion derived from post-mortem cytosine deamination tabulated with the mapDamage program. For each panel, the left and right sides represent the base mismatch proportion at the 5' and 3' ends of the reads mapped to the human reference genome, respectively. The x-axis represents the base position within each read. The red and blue lines indicate the proportion of C>T and G>A misincorporations, respectively. G>A misincorporations are high at the 3' end, while a limited amount of C>T are found at the 5' end, as expected from the Ultra II library preparation protocol.

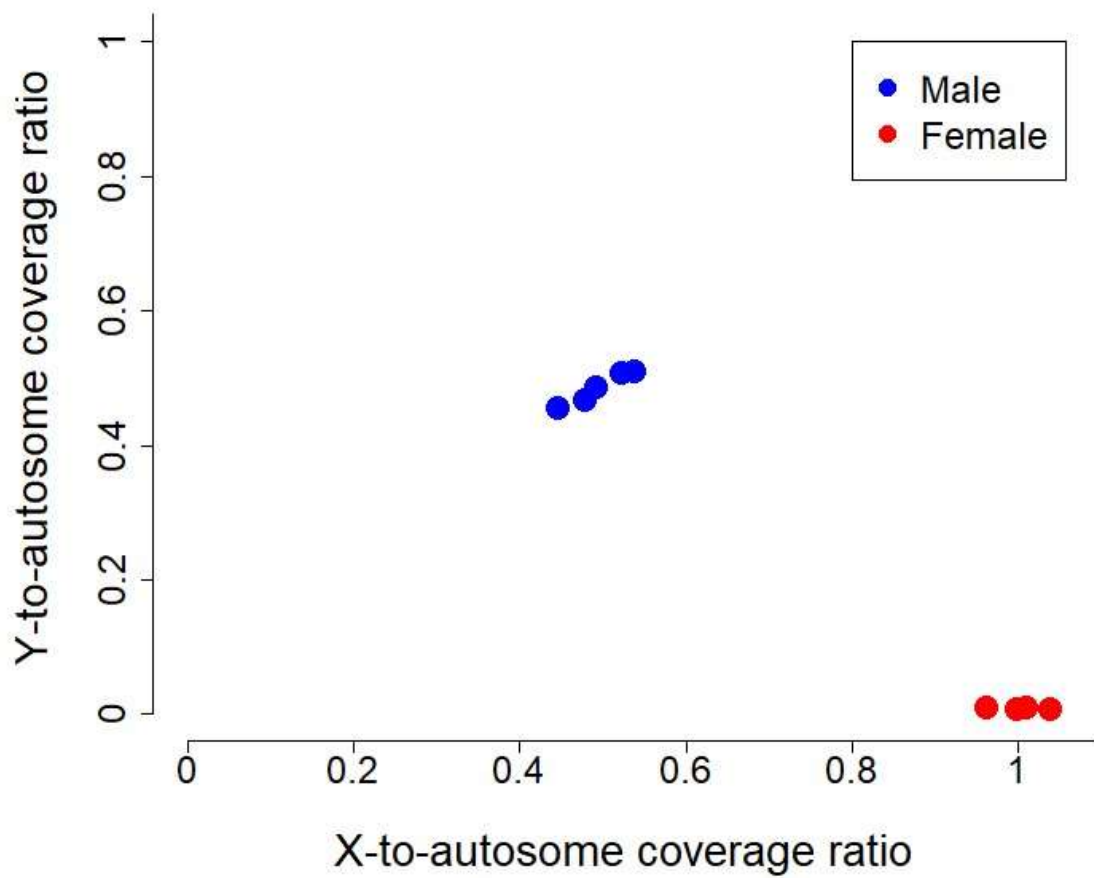


Figure S2. The ratio of the coverage of sex chromosomes to the autosomes. X-axis shows the ratio of X chromosome coverage to autosomal coverage and the y-axis shows the ratio of Y chromosome coverage to autosomal coverage. Two distinct clusters, indicative of the biological sex of each individual, are represented by different colors; blue for males and red for females.

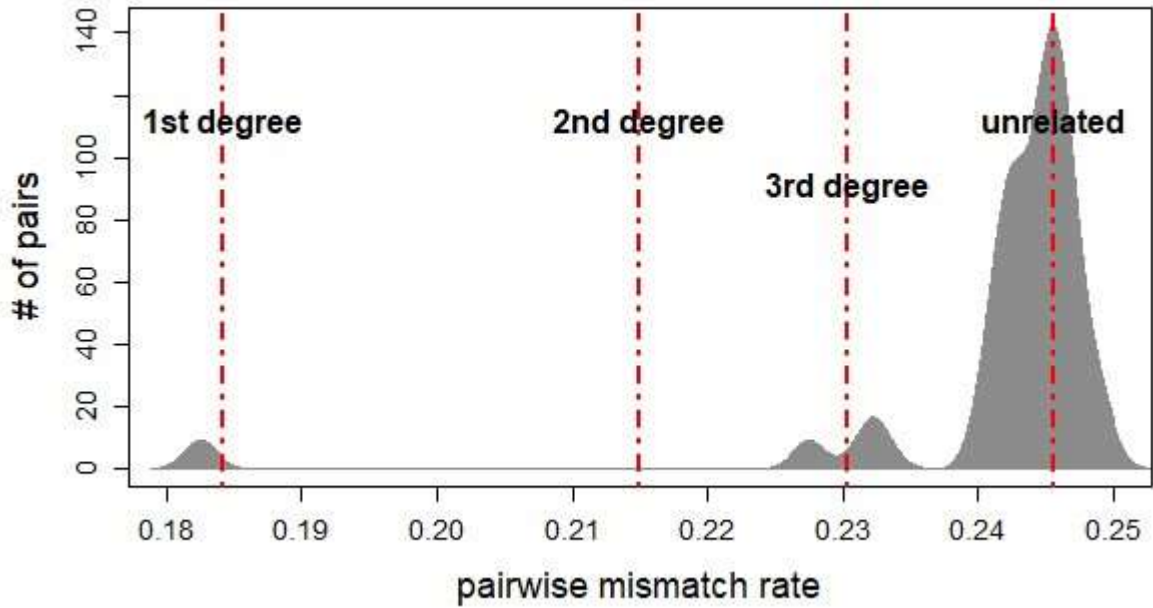
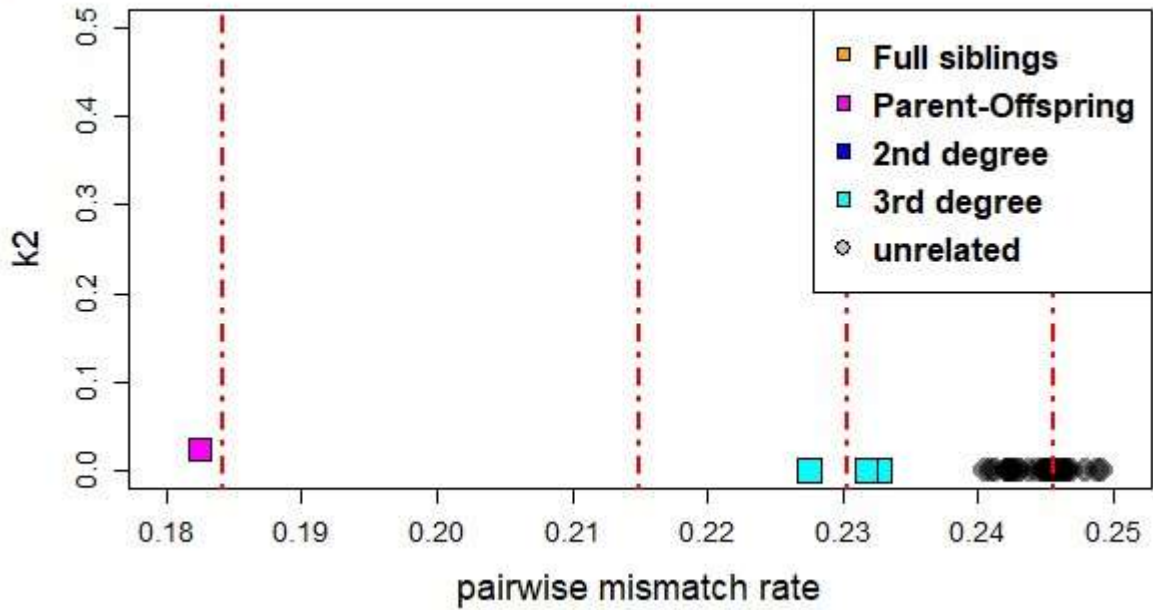
A**B**

Figure S3. Genetic relatives detected within GD and TK sites. To identify genetic relatives, pairwise mismatch rate (pmr) and the probability of sharing two alleles identical by descent (k_2) are estimated. (A) The density of pmr values of every pair of GD and TK individuals. The pmr density was estimated using the R function ‘density’ of R package stats (v4.1.2; The package ‘stats’ is part of R). The red lines indicate the first, second, third-degree relatives and the unrelated from left to right, corresponding to $3/4$, $7/8$, $15/16$, and $16/16$ of the baseline, the pmr values among the unrelated individuals. (B) k_2 values of every pair of GD and TK individuals estimated using lcMLkin. Genetic relationships are shown in different colors (full siblings, orange; parent-offspring, pink; second degree, blue; third degree, light blue; unrelated, gray).

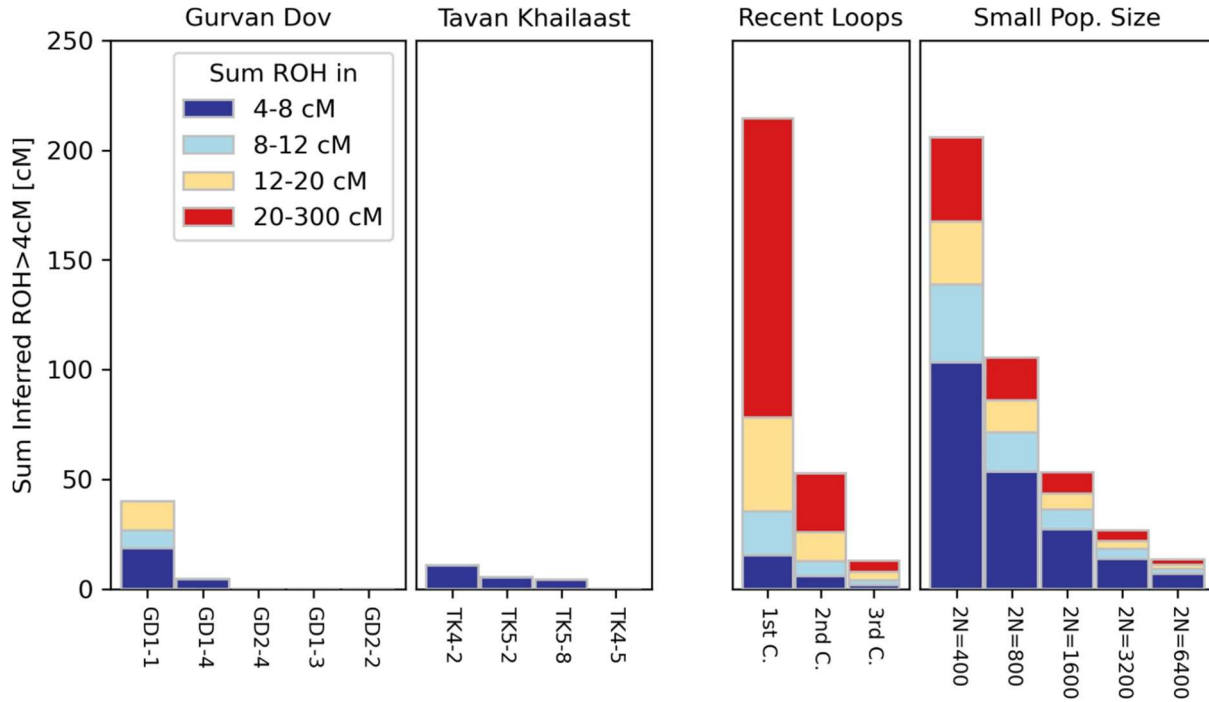


Figure S5. HapROH results for GD/TK individuals. The left panel displays ROH blocks detected in GD/TK individuals. The ROH blocks longer than 20 centimorgan (cM), shown in red, are usually considered as the signature of the close-kin marriage. Short ROH blocks are considered as a signature of a small population size. Expected ROH blocks for close-kin marriage (Recent Loops) and small population size (Small Pop. Size) are shown in the right panel. It illustrates expected ROH blocks resulted from the marriage between the first cousins (1st C.), second cousins (2nd C.), and third cousins (3rd C.) and the small population size with 400 ($2N=400$), 800 ($2N=800$), 1,600 ($2N=1600$), 3,200 ($2N=3200$) and 6,400 ($2N=6400$) individuals. No individual shows long ROH tracts (>20 cM; red), therefore suggesting no recent consanguinity in their ancestors.

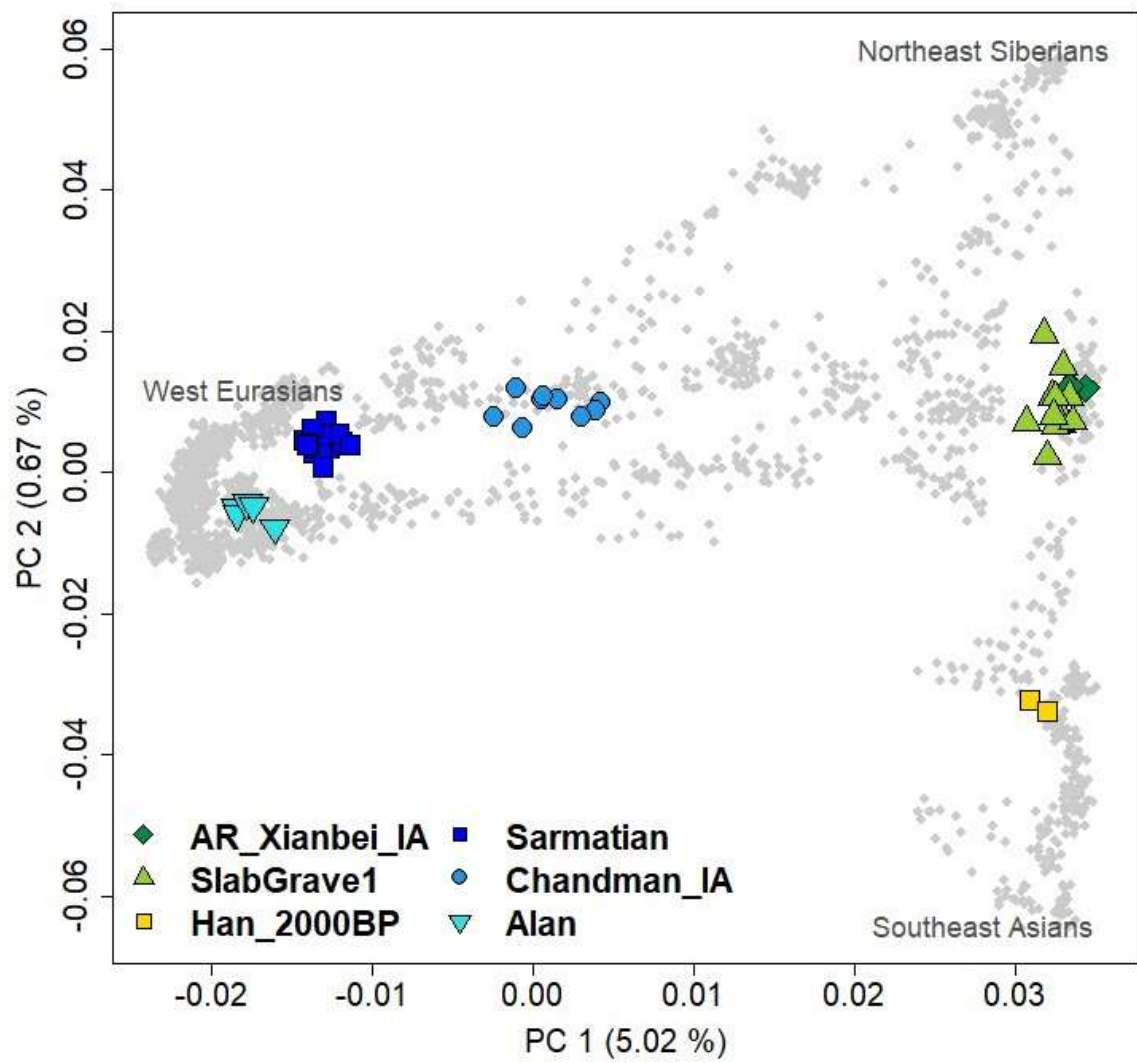


Figure S6. The genetic profiles of the 6 source populations for GD/TK individuals along the two Eurasian PCs. Ancient individuals are projected onto the top two PCs calculated on 2,077 present-day Eurasian individuals (small grey dots). Six ancient populations used as source populations of GD/TK individuals for qpAdm analysis are plotted with the bigger color-filled symbols.

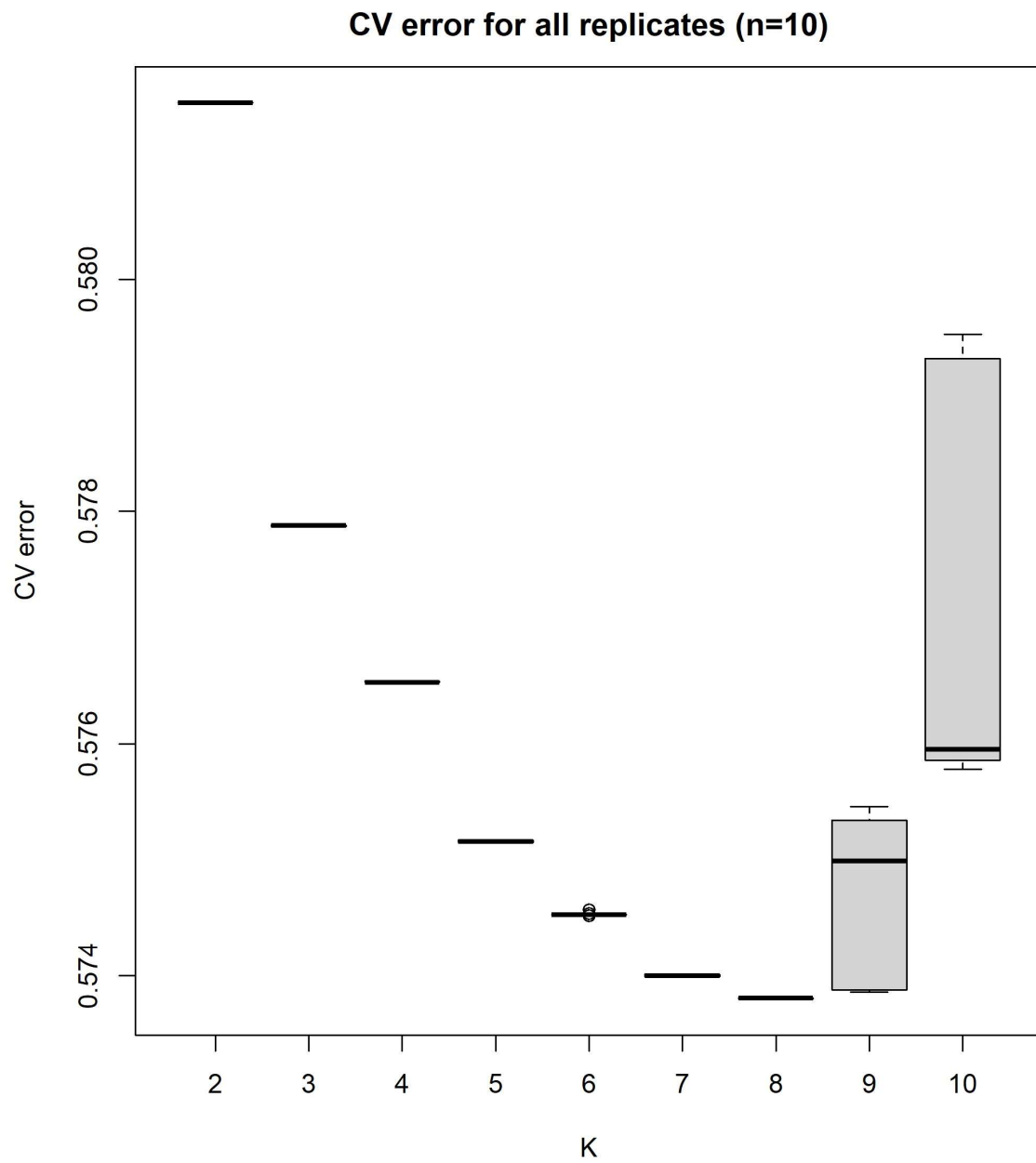
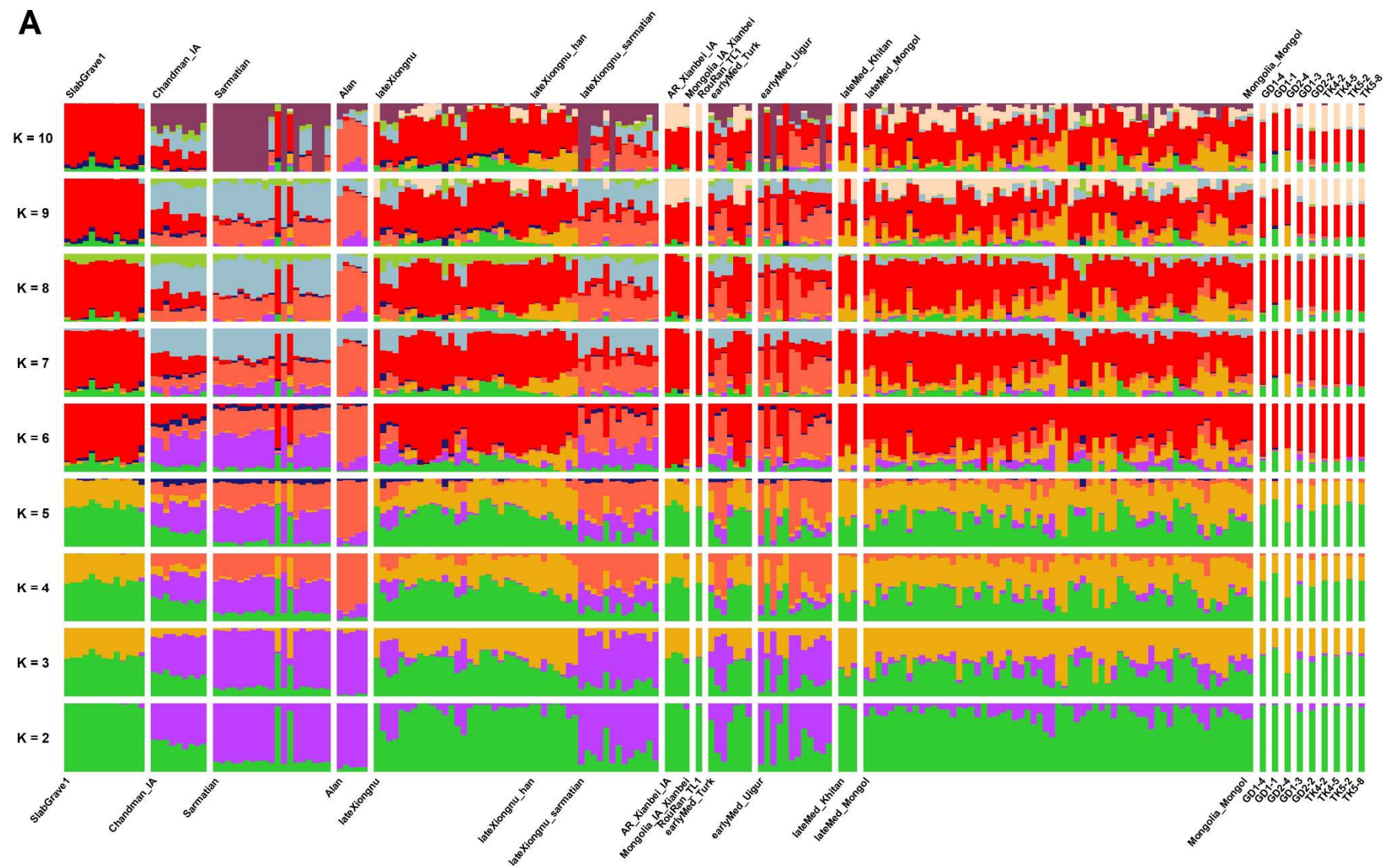


Figure S7. ADMIXTURE cross-validation error analysis for K (K=2-10) with ten replicates. Admixture with K=8 showed the lowest cross-validation (CV) error.

A



B

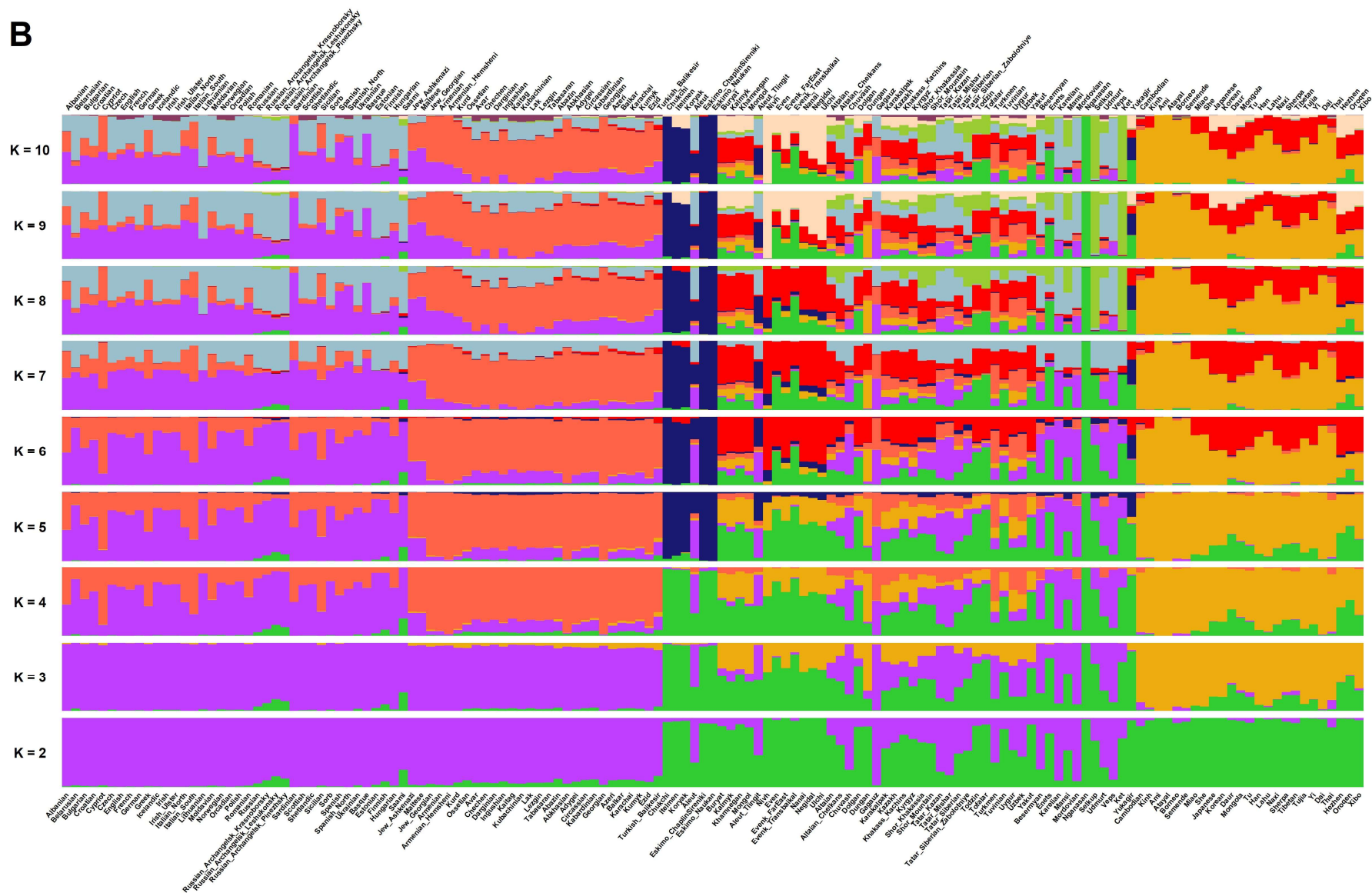


Figure S8. ADMIXTURE results for 1,774 present-day Eurasian and 207 ancient individuals for K=2 to K=10. We report ancestry profiles of ancient (A) and present-day (B) individuals in the form of stacked vertical bar plot. We performed ADMIXTURE for K=2 to K=10 using HumanOrigins dataset. We repeated the analysis for each K for 10 times and plotted the results with the highest log likelihood value. GD and TK individuals show similar genetic profiles with the previously published ancient individuals from contemporaneous time periods. The results for the 1,774 present-day individuals are summarized by taking average ancestry proportion for each population.

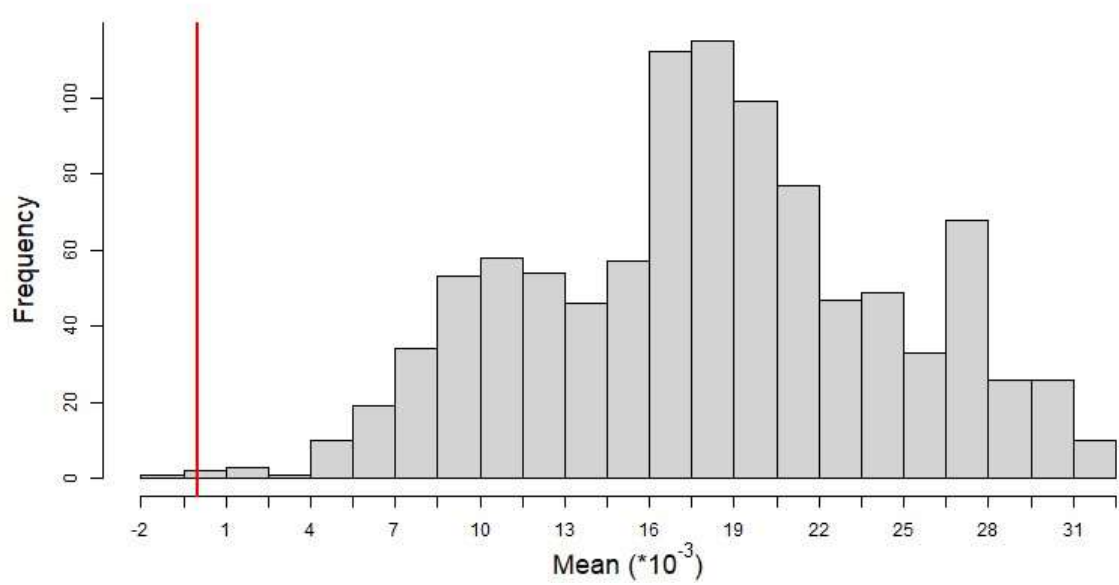


Figure S9. The distribution of the mean values of PC1s of four randomly sampled individuals with identical time period composition to GD. We randomly selected two individuals from Turkic, one individual from Uyghur, and one individual from Khitan (instead of Zubu) period to match the time period composition of four GD individuals. The histogram shows the distribution of the mean PC1 value for 1,000 random sets. The red line indicates the observed mean PC1 value of GD individuals.

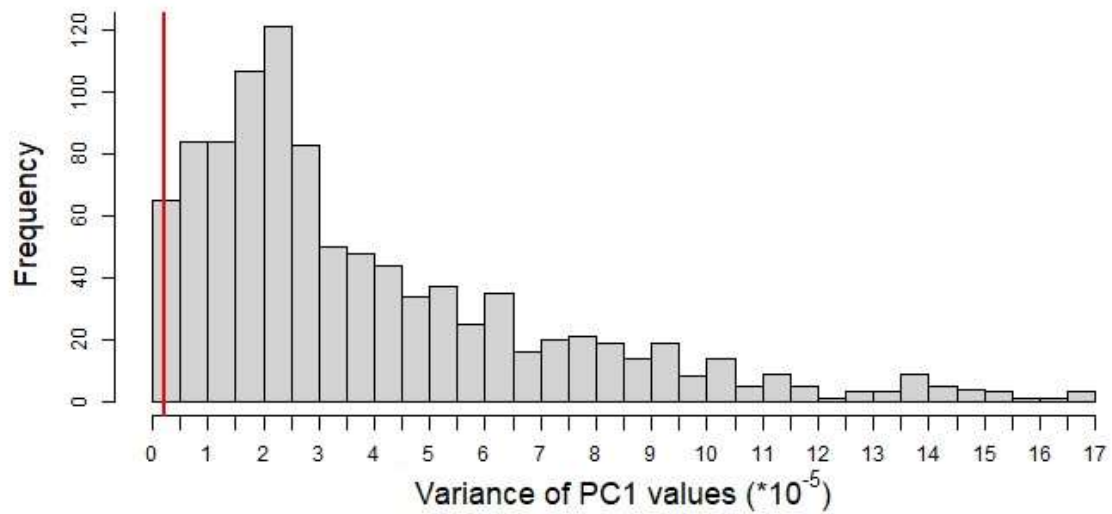
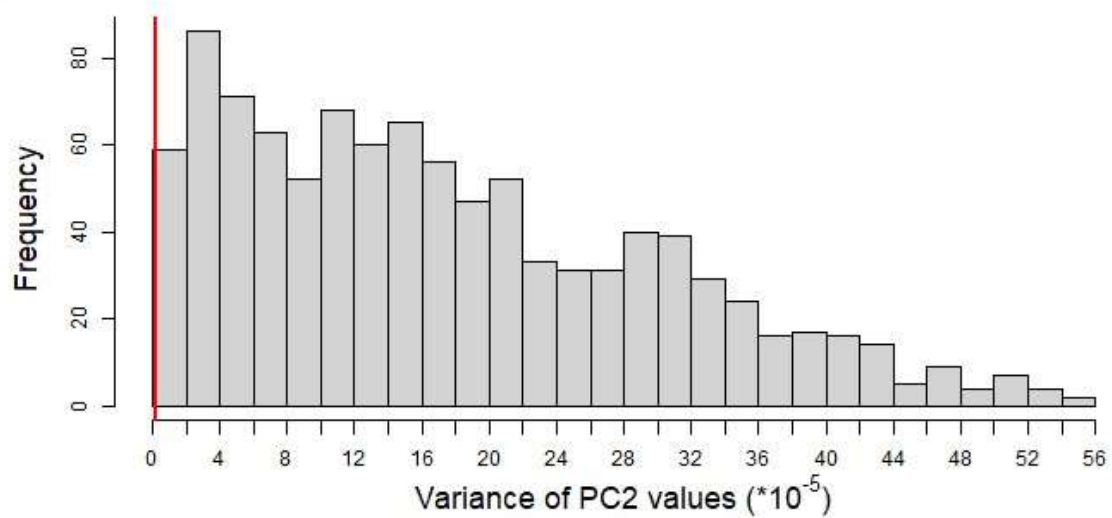
A**B**

Figure S10. Low heterogeneity of the genetic profiles of the four Mongol-period TK individuals. We randomly selected 1,000 sets of four Mongol-period individuals from a combined data set of published and newly reported individuals and calculated the variance in their PC1 and PC2 coordinates. The histograms represent the distribution of variance for (A) PC1 and (B) PC2 among the 1,000 matched random sets. The red lines indicate the observed variance of four Mongol-period TK individuals.