



Mitochondrial genomes uncover the maternal history of the Pamir populations

Min-Sheng Peng¹ · Weifang Xu^{2,3} · Jiao-Jiao Song^{1,4} · Xing Chen¹ · Xierzhatjiang Sulaiman⁵ · Liuhong Cai⁶ · He-Qun Liu¹ · Shi-Fang Wu¹ · Yun Gao¹ · Najmudinov Tojiddin Abdulloevich⁷ · Manilova Elena Afanasevna⁷ · Khudoidodov Behruz Ibrohimovich⁷ · Xi Chen^{8,9} · Wei-Kang Yang^{8,9} · Miao Wu^{8,9} · Gui-Mei Li¹⁰ · Xing-Yan Yang¹¹ · Allah Rakha^{12,13} · Yong-Gang Yao^{12,14,15} · Halmurat Upur² · Ya-Ping Zhang^{1,11,14,15}

Received: 15 April 2017 / Revised: 8 September 2017 / Accepted: 6 October 2017
© European Society of Human Genetics 2017

Abstract

The Pamirs, among the world's highest mountains in Central Asia, are one of homelands with the most extreme high altitude for several ethnic groups. The settlement history of modern humans on the Pamirs remains still opaque. Herein, we have sequenced the mitochondrial DNA (mtDNA) genomes of 382 individuals belonging to eight populations from the Pamirs and the surrounding lowlands in Central Asia. We construct the Central Asian (including both highlanders and lowlanders) mtDNA haplogroup tree at the highest resolution. All the matrilineal components are assigned into the defined mtDNA haplogroups in East and West Eurasians. No basal lineages that directly emanate from the Eurasian founder macrohaplogroups M, N, and R are found. Our data support the origin of Central Asian being the result of East–West Eurasian admixture. The coalescence ages for more than 93% mtDNA lineages in Central Asians are dated after the last glacial maximum (LGM). The post-LGM and/or later dispersals/admixtures play dominant roles in shaping the maternal gene pool of Central Asians. More importantly, our analyses reveal the mtDNA heterogeneity in the Pamir highlanders, not only between the Turkic Kyrgyz and the Indo-European Tajik groups, but also among three highland Tajiks. No evidence supports positive selection or relaxation of selective constraints in the mtDNAs of highlanders as compared to that of lowlanders. Our results suggest a complex history for the peopling of Pamirs by multiple waves of migrations from various genetic resources during different time scales.

Introduction

The settlements of high-altitude regions are great achievements for various human populations in their dispersal history [1]. In Central Asia, the Pamirs are a mountain knot

(i.e., the Pamirian Knot) in the heartland of Asia at the junction of the Himalayas with the Hindu Kush, Tien Shan, Karakoram, and Kunlun ranges (Fig. 1). Due to having high average elevation (2500–4500 m), the Pamirs was first referred to as “Roof of the World” in 1838 [2]. The harsh conditions (e.g., hypoxia, strong ultraviolet radiation, and cold temperature) associated with high altitude makes the Pamirs as one of the most challenging environments for human colonization [3, 4]. The modern Pamir highlanders mainly refer to two linguistic groups. The Indo-European speakers, highland Tajiks, are proposed to be indigenous to the Pamirs [4]. The Turkic-speaking Kyrgyz people are recent immigrants. Their settlement in Pamirs can be traced back to the Turko-Mongol expansions of the medieval period. Some tribes were driven into the Pamirs during the war against the Dzungar Khanate around 1680–1740s AD [5].

The early peopling of the Pamirs was around 10 kya (thousand years ago), as revealed by limited archeological

Min-Sheng Peng, Weifang Xu, Jiao-Jiao Song, and Xing Chen contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-017-0028-8>) contains supplementary material, which is available to authorized users.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

✉ Halmurat Upur
halmurat@263.net

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

Extended author information available on the last page of the article

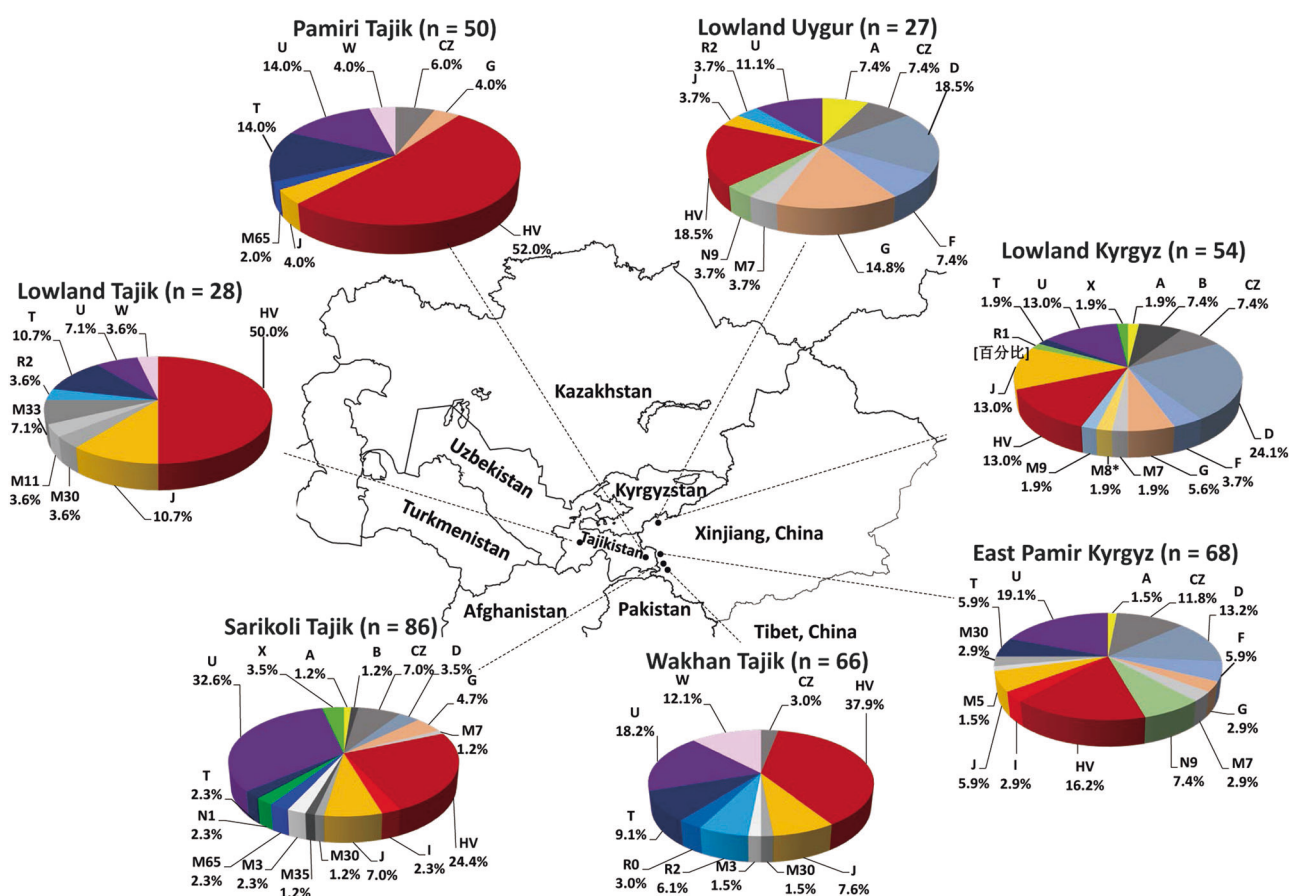


Fig. 1 The haplogroup profiles of populations from Central Asia and neighboring regions. The locations of the populations as well as the sample sizes are showing in the map. The information for each of

populations is described in Table 1. For brevity, some haplogroups are merged. The detailed haplogroup profiles are described in Table S6

sites and rock paintings [6]. Around 4000 years before present, the migrations of Andronovo Culture reached the Pamirs. And then, the Scythians (also known as Saka and Sai), the speakers of Indo-Iranian languages, began to dominate this plateau [7]. It indicates a wave of eastward expansion of the Indo-European language family [7]. Although the Scythians are evident in archeological and historical records [6], their population history remains unclear. Nowadays, the Pamirs are the homelands to the highland Tajiks [4]. Both linguistic and physical anthropological evidence suggest the highland Tajiks likely being the descendants of ancient Scythians [7, 8]. Thus, to investigate the gene pool of modern highland, Tajiks may provide us an opportunity to retrace the origin of early highlanders as well as their settlement history in the Pamirs.

During the past two decades, the study of mitochondrial DNA (mtDNA) has extensively improved our understanding of genetic diversity and population history of Central Asians [9–11]. Several efforts have been done to retrieving the genetic information from ancient mtDNA sequences [12, 13]. However, due to various kinds of difficulty in sample collection, only limited samples of

highland Tajiks were referred recently [14]. Meanwhile, most studies relied upon partial mtDNA sequences (i.e., D-loop and certain coding regions). The limited information hampers further dissection of mtDNA lineages into fine resolution. The recurrent variants occurring in the D-loop can make the haplogroup assignment ambiguous [15]. In previous studies, researchers just selected particular Central Asian individuals of interest for complete mtDNA genome sequencing (e.g., ref. [16]). Such biased sampling and limited sequence data are unsuitable for many demographic analyses [17]. To date, there were few reports focusing on the mtDNA genomes of Central Asians. Lacking of a well-resolved phylogeny of Central Asian mtDNA lineages impedes our knowledge of human evolution and demographic history in this region.

In this study, we focus on the highlanders living in the Pamirs, including three highland Tajik populations and two highland Kyrgyz populations (Table 1). For comparisons, we also investigated three lowland populations neighboring to the Pamirs. We used the next-generation sequencing (NGS) technique to generate the unbiased, population-based data including a total of 382 mtDNA genomes. Our results

Table 1 Summary of populations investigated in this study

Population	Code	Category	Size	Language	Location	Reference
Sarikoli Tajik	ST	Highlander	86	Indo-European	Taxkorgan, Xinjiang, China	This study
Wakhi Tajik	WT	Highlander	66	Indo-European	Taxkorgan, Xinjiang, China	This study
Pamiri Tajik	PT	Highlander	50	Indo-European	Gorno-Badakhshan, Tajikistan	This study
Lowland Tajik	LT	Lowlander	28	Indo-European	Dushanbe, Tajikistan	This study
East Pamir Kyrgyz	EPK	Highlander	68	Turkic	Taxkorgan, Xinjiang, China	This study
West Pamir Kyrgyz	WPK	Highlander	3	Turkic	Gorno-Badakhshan, Tajikistan	This study
Lowland Kyrgyz	LK	Lowlander	54	Turkic	Artux, Xinjiang, China	This study
Lowland Uygur	LU	Lowlander	27	Turkic	Artux, Xinjiang, China	This study
Sherpa	She	Highlander	76	Tibeto-Burman	Shigatse, Tibet, China	48
Beijing Han Chinese	CHB	Lowlander	103	Sino-Tibetan	Beijing, China	1000 Genomes Project
Punjabi	PJL	Lowlander	96	Indo-European	Lahore, Pakistan	1000 Genomes Project
Persian	Per	Lowlander	181	Indo-European	Iran	18
Qashqai	Qas	Lowlander	112	Turkic	Iran	18
Azeri	Aze	Lowlander	30	Turkic	Baku, Azerbaijan	17
Armenian	Arm	Lowlander	30	Indo-European	Yerevan, Armenia	17
Georgian	Geo	Lowlander	28	Kartvelian	Batumi, Georgia	17

Note: CHB and PJL are retrieved from the 1000 Genomes Project (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/MT/chrMT_sequences_2534.20160505.fasta.gz)

not only help further understand the mtDNA phylogeny in Central Asia, but also shed novel light on the peopling of the Pamirs.

Materials and methods

Sampling and data collection

We collected the peripheral blood samples of 382 unrelated adult individuals, which can be attributed to eight populations based on geographic location and ethnic information (Table 1 and Fig. 1). Written informed consents were obtained from all individuals. The study protocol was approved by the Internal Review Board of Kunming Institute of Zoology, Chinese Academy of Sciences, Xinjiang Medical University, and E.N. Pavlovsky Institute of Zoology and Parasitology, Academy of Sciences of Republic of Tajikistan. The published population data from East Asia, South Asia, West Asia [18], and Caucasus [17] were collected for comparison (Table 1).

DNA extraction, mitochondrial genome amplification, and sequencing

Genomic DNA was extracted from the blood samples by using the phenol-chloroform method. The complete mtDNA genome was amplified by the long-range PCR protocol to generate two overlapping amplicons of ~8.5 kb [19]. We constructed the pooled libraries according to the instruction

provided by Ion Torrent (Life Technologies). The pooled libraries were amplified by emulsion PCR and sequenced in the Ion 316™ Chip with the Ion PGM™ 400 Sequencing Kit. The raw data have been deposited into Genome Sequence Archive (<http://bigd.big.ac.cn/gsa>; project ID: PRJCA000486; accession numbers SAMC014618-SAMC014999; Table S1).

NGS data analyses

In the initial quality control, Torrent Suite removes the low-quality base to make sure the average Q value is more than 15 within 30 bp window on the 3' end of reads. We used the revised Cambridge Reference Sequence (rCRS; GenBank: NC_012920) [20] as the reference. The reads mapping and variants calling were carried out with Ion Torrent Software Suite Plugin Variant Caller v5.0.2.1 as recommended by the manufacturer. For all the scored variants in vcf files, we further manually checked the corresponding bam files with Integrative Genomics Viewer [21] to confirm the status. We developed a python script to transfer the confirmed vcf file (vcf format) into the consensus sequence (fasta format) with the cutoff value of heteroplasmy as 0.2. The sequences of fasta format are available at GenBank under accession numbers MF522841-MF523222.

Analyses of mtDNA genome sequences

By using HaploGrep 2 [22], we scored the variants out for each of sequences relative to rCRS, and then assigned the

mtDNA sequences into specific haplogroups. The results were further confirmed by using MitoTool [23]. The transition at 16519, the C-length variants in regions 16180–16193 and 303–315, and AC indels at 515–522 were disregarded in subsequent analyses. The nomenclatures were referred to PhyloTree (<http://phylotree.org/>; Build 17) [24]. The mtDNA haplogroup trees were reconstructed manually.

The indexes for nucleotide and haplotype diversity, and Tajima's D, Fu's Fs, and AMOVA tests were computed with ARLEQUIN 3.5.1.3 [25]. We compared the numbers of synonymous (S) and nonsynonymous (N) substitutions in the mtDNA protein-coding regions from the Pamir highlanders and the surrounding lowlanders to estimate the potential selective constraints as described before [26]. The 3639 sequences after a stringent quality-control procedure [26] were retrieved as the general populations for comparisons.

We estimated the average sequence divergence of the haplotypes to their most recent common ancestor by the rho statistic [27]. A heuristic estimate of the standard error (sigma) was calculated as described before [28]. The rho \pm sigma was converted into the coalescent age for certain haplogroup by using the calibration rates for mtDNA coding region (np 577–16024) variants [29] and only synonymous substitutions [30], respectively. We also employed BEAST v1.8 [31] to date the nodes based on 382 coding region sequences with the fixed rate as 1.404×10^{-8} substitutions per site per year [29]. The strict clock model and coalescent Bayesian skyline were adopted. We used jModelTest 2.1.3 [32] to select GTR + I + G substitution model as well as estimate the related parameters. MCMC was run for 100,000,000 generations and sampled every 2000 generations. The burn-in for trees was set as 10%.

The Bayesian skyline plots (BSPs) for each of populations were conducted by using BEAST v1.8 [31] based on coding region sequences. The TrN + I + G substitution model as well as the related parameters were estimated with jModelTest 2.1.3 [32]. We chose strict clock model with the rate 1.404×10^{-8} substitutions per site per year [29]. Each MCMC was run for 300,000,000 generations and sampled every 5000 generations. The effective sample size values for the parameters of interest were over 300. The results were visualized with Tracer v1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>).

The approximate Bayesian computation (ABC) analyses were done by using DIYABC 2.1 [33]. We used Beijing Han Chinese and Persian populations (Table 1) to represent the modern East and West Eurasians, respectively. It should be noted that the two populations cannot be treated as the parental or ancestral proxies for Central Asians. We only considered Kyrgyz and Tajik populations. We carried out the pairwise sample/population comparison in terms of

using four summary statistics: number of segregating sites in the pooled sample, mean of within-sample pairwise differences, mean of between-sample pairwise differences, and F_{ST} between two samples, as suggested before [17]. The number of simulated data sets in computation is 200,000. The scenario prior combination pre-evaluating and model/scenario checking are done by using PCA based on the Euclidian distance between observed and simulated summary statistics.

Results

Sequencing of mtDNA

After filtering the unmapped reads and reads with poor quality, the average mapped reads are 26,492 per individual (minimum 4346; maximum 76,814) (Table S1). The average depth is $300\times$ per individual, ranging from $51\times$ to $964\times$. The mtDNA genome is well covered for each of the individuals. Due to some technological defects of NGS, certain positions had higher false insertion—deletion rates [34] and then identified as phantom variants. We detect a total of 53 positions with potential errors and then corrected them (Table S2 and see “Materials and methods” for details).

MtDNA genetic diversity

We excluded the West Pamir Kyrgyz consisting of only three individuals (Table 1) in the subsequent analyses at the population level. The genetic diversity indexes in seven Central Asian populations are at a similar level, with the values observed in populations from East Asia (Beijing Han Chinese and Sherpa), South Asia (Punjabi), West Asia (Persian and Qashqai), and Caucasus (Armenian, Azeri, and Georgian) (Table 2). In Central Asians, the relatively low nucleotide diversity and haplotype diversity values are observed in lowland Tajik (0.00168) and Wakhi Tajik (0.979), respectively. The values of Tajima's D and Fu's Fs are negative in all populations, indicating that the high frequencies of rare or private variants in the mitochondrial genomes. It indicates that those populations underwent recent expansions [35], although some negative Fu's Fs values are not significant statistically.

MtDNA haplogroup profiles

In terms of PhyloTree (<http://www.phylotree.org/>; Build 17) [24], we classify all 382 mtDNA sequences into 73 specific (sub-)haplogroups within the non-African macro-haplogroups M, N, and R (Table S3). We then incorporate all sequences to build the haplogroup tree for Central

Table 2 Genetic diversity indexes and neutrality tests for population mitochondrial genomes

Population	Size	No. HT	Nucleotide diversity	HT diversity	Tajima's D	Fu's Fs
Sarikoli Tajik	86	67	0.00204 ± 0.00007	0.993 ± 0.003	-2.180**	-17.648**
Wakhi Tajik	66	40	0.00187 ± 0.00009	0.979 ± 0.007	-1.811*	-1.634
Pamiri Tajik	50	38	0.00180 ± 0.00013	0.985 ± 0.008	-1.842**	-4.870
East Pamir Kyrgyz	68	45	0.00204 ± 0.00007	0.986 ± 0.005	-2.021*	-3.214
Lowland Kyrgyz	54	45	0.00213 ± 0.00007	0.994 ± 0.005	-2.187**	-7.868*
Lowland Tajik	28	23	0.00168 ± 0.00013	0.984 ± 0.014	-1.912*	-2.117
Lowland Uygur	27	24	0.00206 ± 0.00011	0.991 ± 0.013	-2.025*	-2.822
Beijing Han Chinese	103	103	0.00211 ± 0.00005	1.000 ± 0.0014	-2.442**	-24.021**
Punjabi	96	87	0.00198 ± 0.00006	0.995 ± 0.004	-2.456**	-24.008**
Qashqai	112	94	0.00192 ± 0.00007	0.996 ± 0.002	-2.439**	-23.940**
Persian	181	163	0.00198 ± 0.00006	0.9985 ± 0.0009	-2.575**	-23.757**
Azeri	30	30	0.00161 ± 0.00011	1.000 ± 0.009	-2.017**	-6.764*
Armenian	30	29	0.00199 ± 0.00012	0.998 ± 0.009	-2.132**	-11.175**
Georgian	28	28	0.00151 ± 0.00013	1.000 ± 0.010	-2.217**	-9.834**

Note: The calculation is based on 16511 bp mtDNA sequences. The variants 309.XC, 315.1C, 16182C, 16183C, 16193.1C(C), 16519, and heteroplasmy sites are not considered

HT haplotype

* $P < 0.05$; ** $P < 0.01$

Asians (Table S4). No basal lineages directly split from macrohaplogroups M, N, and R are defined. Most (sub-)haplogroups in Central Asians are derived from the characterized haplogroups in East and West Eurasians. Some South Asian autochthonous haplogroups, such as M30, M33a, M35, M3, and M5, are also found (Table S3). In the Turkic group, the most dominant haplogroup is D4, accounting 13.2% in East Pamir Kyrgyz, 20.4% in lowland Kyrgyz, and 18.5% in lowland Uygur. In the Indo-European group, haplogroup H is prevailing in Sarikoli Tajik (22.1%), Pamiri Tajik (48.0%), Wakhi Tajik (37.9%), and lowland Tajik (32.1%). Within haplogroups, our data characterize a series of new subhaplogroups. For instance, three sequences (ST143, WT199, and WT217) cluster as a subhaplogroup of haplogroup H (Table S4). Interestingly, subhaplogroup U5a2a1, which was detected in one 700-year-old individual excavated from the eastern Pamirs [16], exists in four modern individuals of four populations (Table S4).

In terms of the haplogroup frequency profiles (Table S3), we carry out the analysis of molecular variance (AMOVA) to discern the genetic differentiations among groups. The division between the Turkic (East Pamir Kyrgyz, lowland Kyrgyz, and lowland Uygur) and Indo-European (lowland Tajik, Pamiri Tajik, Sarikoli Tajik, and Wakhi Tajik) speakers contributes 1.86% of total variation, and it is significant in AMOVA ($P = 0.03128 \pm 0.00588$). We further compare the highlanders and lowlanders within the Turkic and Indo-European groups, respectively. The differentiation is not significant in both East Pamir Kyrgyz vs. lowland Kyrgyz and lowland Uygur ($P = 0.33040 \pm 0.01275$) and Pamiri Tajik and Sarikoli Tajik and Wakhi Tajik vs.

lowland Tajik ($P = 0.33040 \pm 0.01275$) in AMOVA, suggesting a small differentiation of mtDNA haplogroup profiles between the highlanders and lowlanders.

The coalescent time of mtDNA lineages

The Central Asian mtDNA genomes cluster with the published sequences into 252 (sub-)haplogroups (Table S5). Among of them, the numbers of sequences are less than six in 85 subhaplogroups. We adopted the rho statistic [27] to calculate the coalescent ages for mtDNA lineages from Central Asians based on the haplogroup tree (Table S4). When just considering the mean values, the most ancient coalescent ages for Central Asians (including both highlanders and lowlanders) are observed in haplogroup U (Fig. 2 and Table S5), including U2b2 (~32–30 kya), U2c1b (~39–32 kya), and U3b (~32–30 kya). In the Pamir highlanders, the most ancient lineages of subhaplogroups U2b2 and U3b occur in East Pamir Kyrgyz (1/68; 1.4%) and Sarikoli Tajik (1/86, 1.1%), respectively (Table S4). The coalescence ages of lineages of haplogroups B4c1a, D4m, D4v, F2e, M30 (WT292), M33a3b, M7c1, T2g, and $\times 2p$ fall within the LGM. Those sporadic lineages make contributions of 2.9% (11/382) for the maternal gene pool. Both in the Pamir highlanders and the surrounding lowlanders, most (>93%) lineages coalesced after the LGM. More than 70% lineages coalesced within 10 kya. Notably, some subhaplogroups contain the small number of sequences, or show not in the ideally star-like structure. Those dating results are with huge standard errors (Table S5). The Bayesian dating for the 382 mtDNA genomes shows that

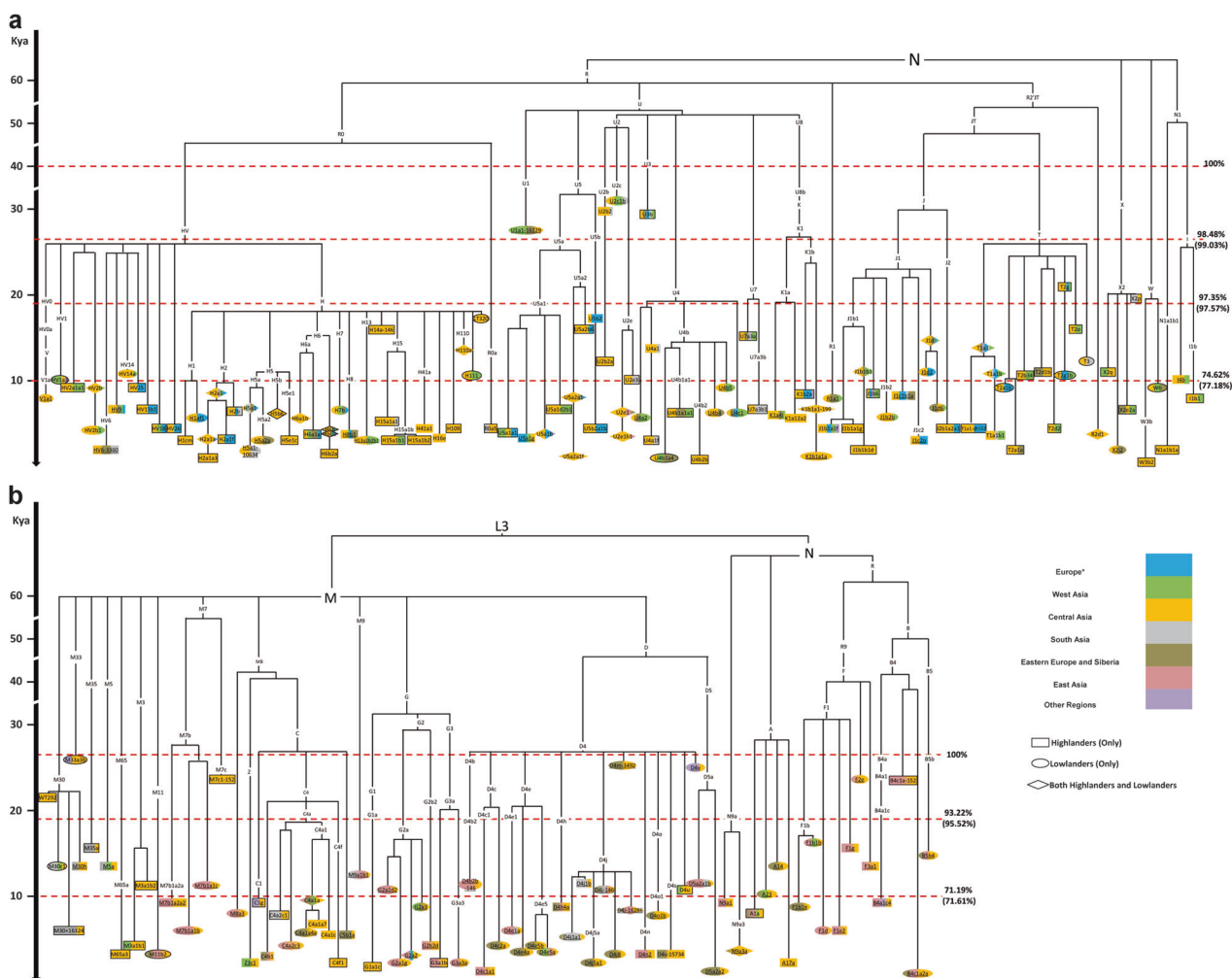


Fig. 2 Schematic trees of mtDNA lineages in Central Asian populations. The trees are scaled with the rho statistic based on coding region variations (Table S5). Colors indicate geographic affiliation of sequences (Table S4). The percentage indicates the proportion in both Pamir highlanders and surrounding lowlanders. The percentage in parentheses indicates the proportion in highlanders only. Haplogroups referring to highlanders, lowlanders, and both are labeled as rectangle, oval, and rhombus, respectively. Haplogroups including sequences

from Indo-European and Turkic speakers are framed and underlined, respectively. WT292 and LT320 can be assigned as M30* and H*, respectively. Their ages are retrieved from the previous study [30]. **a** Schematic trees of macrohaplogroup M and haplogroups N9, A, B, and R9. Most lineages originate from East Asia; and some (i.e., M3, M5, M30, and M65) are from South Asia. **b** Schematic trees of haplogroups N1, W, X, R0, R2, JT, and U. These lineages are prevailing in West Eurasia and South Asia

more than 95% sequences coalesced after the LGM (Fig. S1), which is largely in agreement with the results of rho statistic (Fig. 2).

Genetic relationship of the Pamir highlanders with other populations

We employ PCA based on the distribution frequency of 23 haplogroups (Table S6) to display the relationships of the Pamir highlanders with other Eurasian populations (Fig. 3). The first principal component (PC1) mainly reflects a separation between East and West Eurasian. The East Pamir Kyrgyz, lowland Kyrgyz, and lowland Uyghur populations are close to Beijing Han Chinese. The lowland Tajik, Pamiri

Tajik, Sarikoli Tajik, and Wakhi Tajik cluster with the populations from Iran (Persian and Qashqai) and Caucasus (Armenian, Azeri, and Georgian). The PC2 further distinguishes the genetic differentiation within the Indo-European Tajiks. It separates Sarikoli Tajik and Pamiri Tajik from lowland Tajik and Wakhi Tajik. In the east pole, the Turkic speakers (East Pamir Kyrgyz, lowland Kyrgyz, and lowland Uyghur) are close to each other.

Inference of maternal effective population size changes

We conduct Bayesian skyline plot (BSP) to infer historical variations in maternal effective population size based on mtDNA coding region sequences (Fig. 4). The earliest

population increase took place 45–50 kya, which was most likely to be attributed to the common background of ancient

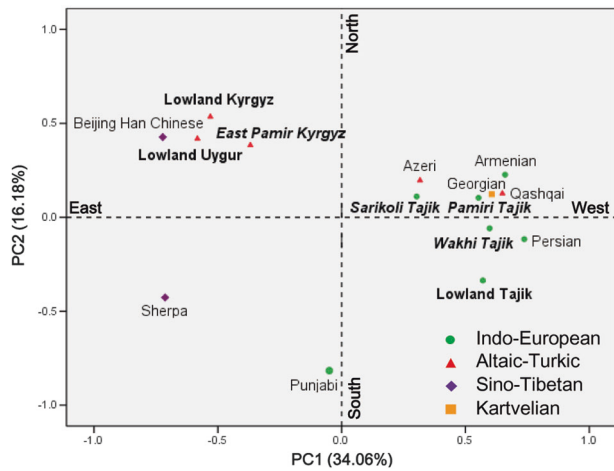


Fig. 3 PCA of populations from Central Asia and neighboring regions. PC map of populations based on mtDNA haplogroup frequencies (Table S5). The populations generated in this study are in bold

demographics in various Asian populations [36]. From 40 to 20 kya, population growth lasts in lowland Uyghur, whereas constant size is observed in other six populations. After the LGM (26.5–19 kya) [37] but before the Holocene, the population growths occur in East Pamir Kyrgyz, lowland Kyrgyz, lowland Tajik, Pamiri Tajik, and Sarikoli Tajik populations but not in lowland Uyghur and Wakhi Tajik populations. During the last 2 kya, a decline of population size is seen in all seven Central Asian populations as well as in the Iranian population. The recent declines in BSPs are in conflicts with the expansions revealed by the negative values of Tajima's D and Fu's Fs (Table 2). One possible explanation for this discrepancy is that the false signals of population decline can be detected in BSP with biologically plausible combinations of population structure and sampling strategy [38].

Approximate Bayesian computation analyses

To address the origin and admixture history of the Pamir highlanders, we perform ABC to test different scenarios.

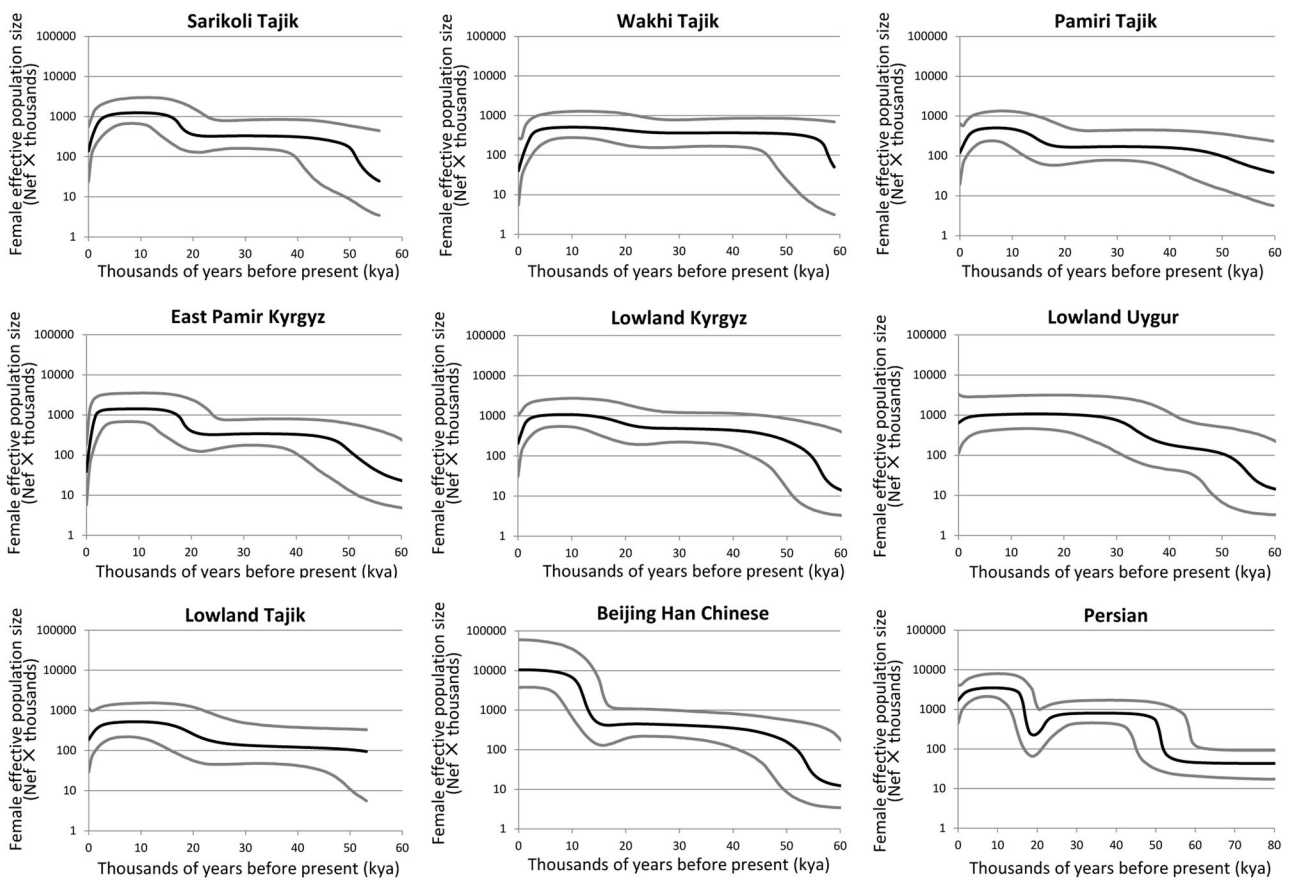


Fig. 4 BSP of female effective population size (N_{ef}) through time for nine populations. The bold black line represents the median population size estimate from Bayesian posterior distribution. The gray lines delimit the 95% highest posterior density (HPD) boundaries,

accounting for uncertainty in the reconstructed phylogeny and substitution model parameters. All the lines were drawn within the lower 95% HPD boundary of the maximum time to the coalescent age

We propose the scenario of three populations: West Eurasian—Pamir highlander—East Eurasian (Fig. 5a). The admixture model indicates genetic admixture playing a fundamental role in the origin of Pamir highlander. Whereas the split model suggests that the Pamir highlander is an isolated population receiving minor or few gene flows from West and East Eurasians. In the more complex scenario of four populations: West Eurasian—Ancient Central Asian—Pamir highlander—East Eurasian, we propose the Ancient Central Asian as the unsampled ancestor population. And our aim is to test the admixture-admixture and split-admixture models for the origin of Pamir highlander (Fig. 5b).

The pre-evaluation based on PCA shows the plots of observed data sets are overlapping with those of scenario prior combinations (Fig. S2), suggesting our proposed scenarios and priors are compatible with the observed data sets. In the scenarios of three populations, the data of East Pamir Kyrgyz, lowland Kyrgyz, and Sarikoli Tajik fit the admixture model well, while lowland Tajik, Pamiri Tajik, and Wakhi Tajik prefer the split model (Fig. S3). When considering the more complex scenarios of four populations, with the exception of lowland Kyrgyz fitting, the admixture-admixture model, the rest five populations prefer the split-admixture model. The model checking based on PCA suggests the selected models with posterior parameters estimated by ABC fit the observed data sets better as compared with those models with prior parameters (Fig. S4). The posterior parameters of interests are shown in Table 3. The parameter estimation for the same population varies in different scenarios. For instance, the split of Pamiri Tajik from West Eurasian is dated 405 and 1190 generations according to the scenarios of three and four populations, respectively. It should be noted that, in the scenario of four populations, the recent admixture time (both mean and median values) is older than the ancient split time in lowland Tajik. It implies the current model is unable to distinguish the ancient split and recent admixture. Or both demographic events occurred simultaneously, so the time scales were overlapping. The most ancient split time is observed in Wakhi Tajik (1410–1510 generations, mean value) around 28–30 kya (a generation time of 20 years), which is before the LGM. In the Pamir highlanders, the admixture time (not the split time) is estimated within 15 ky, that is after the LGM. The gene flow from East Eurasian is stronger in East Pamir Kyrgyz (35.3–55.8%) than Pamiri Tajik (17.6%), Sarikoli Tajik (19.7–21.8%), and Wakhi Tajik (16.4%).

Detection of selection

The number of nonsynonymous substitution (N), the number of synonymous substitution (S), and the ratio N/S in the

Pamir highlanders, the lowlanders, and the general population [26] are shown Fig. S5. The N/S ratio in the highlanders (0.430) is higher than that in the lowlanders (0.381), but lower than that in the general population (0.495). The ratio differences between the highlanders and the lowlanders/the general population are not significant ($P > 0.05$; Table S3). The distribution of N/S ratio in 13 protein-coding genes indicates no significant differences between the highlanders and lowlanders with an exception of *COX2* gene (Table S3). In *COX2* gene, the N/S ratio in the highlanders (2/27) is significantly lower ($P = 0.013$) than that in the lowlanders (82/213). According to the genetic affinity revealed by PCA (Fig. 3), we further compared the highland Tajiks and highland Kyrgyz with their corresponding lowlanders (Table S7). Only the N/S ratio of *CYTB* gene in Pamiri Tajik (17/12) is significantly higher ($P = 0.016$) than that (33/69) in the corresponding lowlanders. Nevertheless, the P value is higher than 0.05 after the Bonferroni correction.

Discussion

Central Asia, located in the Eurasian heartland, is a hotspot for studying human evolutionary history. The paleoanthropological and archeological evidence suggests that Central Asia was occupied by ancient hominin since the early Pleistocene (800–900 ka) [39]. It was also inhabited by Neanderthals [40]. No archaic mtDNA lineages with origins from the other archaic hominids in Central Asia, such as Denisovans [41] and Neanderthals [40], are detected (Fig. 1 and Tables S3 and S4). Given that some gene flows from archaic hominids are revealed in various populations around Central Asia (e.g., ref. [42]), the lack of mtDNA introgression can be explained by the small effective size of mtDNA (i.e., one quarter that of the autosomes), which may be subjected to severe bottleneck effect/genetic drift. There is a possibility that the current absence of archaic hominid types may be caused by insufficient sampling. It is also compatible with the model of strong reproductive isolation between modern humans and archaic hominids [43].

We first provide a detailed mtDNA haplogroup tree at the highest molecular resolution for Central Asian populations (Table S4). As compared with previous studies mainly based on mtDNA D-loop sequences and certain coding regions [9–11], it can provide deeper insights into the origins of Central Asians. Especially in terms of the mtDNA genomes, we can date the Central Asian lineages into a fine scale. We find no basal lineages of macrohaplogroups M, N, and R in our data. The mtDNA lineages in Central Asians are classified into a subset of those typical of the East and the West Eurasians rather than ancient and/or unique lineages. The estimated ages for all lineages are within 40

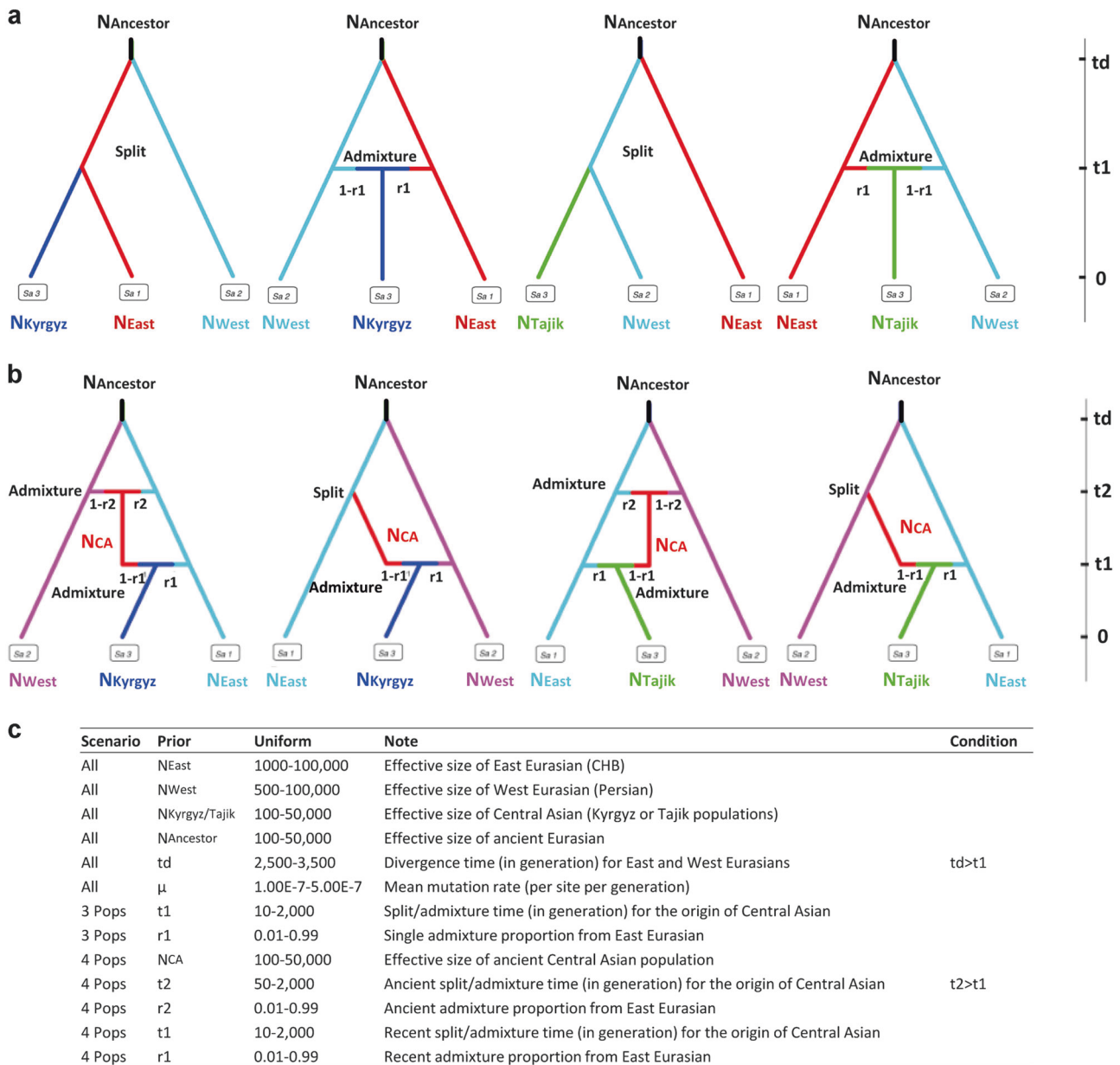


Fig. 5 Approximate Bayesian computation analyses. **a** The scenario of three populations. The split and admixture models for Kyrgyz (East Pamir Kyrgyz and lowland Kyrgyz) and Tajik (lowland Tajik, Pamiri Tajik, Sarikoli Tajik, and Wakhi Tajik) are included. **b** The scenario of four populations. The split-admixture and admixture-admixture

models for Kyrgyz and Tajik are included. **c** The priors are list in the table. The generation time is 20 years [50]. The individual locus mutation rate ranges 1.00×10^{-9} to 5.00×10^{-6} substitutions per site per generation, following a gamma distribution (Shape = 2)

kya (Fig. 2 and Table S5), i.e., after the initial arrival of modern human in Central Asia during 50–40 kya [6]. It supports the genetic admixture hypothesis (e.g., refs. [9, 10]) rather than the Central Asian heartland hypothesis which proposed that Central Asia was a heartland and reservoir for the genetic diversity of Eurasia and America [44]. Also, our results do not support the scenario that people carrying mtDNA N lineages crossed Asia via a northern route and then migrated to Australia [45]. Most (i.e., 93–97%) lineages coalesced after the LGM (Fig. 2 and

Fig. S1). As a result, the post-LGM and/or later dispersals/admixtures play dominant roles in shaping the maternal gene pool of Central Asians.

According to the population mtDNA genomes, we are able to conduct extensive analyses to uncover the population history for Pamir highlanders. The genetic differentiation between the Tajik and Kyrgyz groups revealed in PCA (Fig. 3) reflects their different linguistic sources. Both East Pamir Kyrgyz and low Kyrgyz populations have close relationships in PCA and similar patterns in BSPs,

Table 3 Summary of posterior estimates of parameters of the scenarios fitted to the mtDNA data

Scenario 3 Pops Preferred Model Parameter	East Pamir Kyrgyz		Lowland Kyrgyz		Lowland Tajik		Pamiri Tajik		Sarikoli Tajik		Wakhi Tajik	
	Admixture		Admixture		Split		Split		Admixture		Split	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
NEast	6.91E+04	7.00E+04	5.98E+04	5.84E+04	4.37E+04	4.03E+04	4.22E+04	3.95E+04	6.59E+04	6.58E+04	5.91E+04	5.83E+04
NWest	7.92E+04	8.04E+04	6.56E+04	6.52E+04	7.60E+04	7.70E+04	8.08E+04	8.12E+04	8.67E+04	8.76E+04	6.45E+04	6.37E+04
NTarget	2.18E+04	2.05E+04	2.87E+04	2.86E+04	2.83E+04	2.79E+04	1.69E+04	1.51E+04	3.01E+04	3.05E+04	3.69E+04	3.74E+04
t ^{1a}	3.06E+02	2.18E+02	3.13E+02	2.01E+02	4.05E+02	2.92E+02	3.07E+02	2.11E+02	4.59E+02	3.87E+02	1.51E+03	1.58E+03
r1	5.58E-01	5.80E-01	7.66E-01	8.15E-01	na	na	na	na	2.18E-01	1.82E-01	na	na
td ^a	3.00E+03	2.99E+03	2.86E+03	2.80E+03	2.96E+03	2.94E+03	3.07E+03	3.09E+03	2.89E+03	2.84E+03	3.04E+03	3.05E+03
NAncesor	7.87E+02	5.91E+02	7.12E+02	5.61E+02	4.45E+02	3.41E+02	5.12E+02	4.15E+02	9.83E+02	7.01E+02	4.18E+02	3.20E+02
μ ^b	3.28E-07	3.25E-07	3.71E-07	3.70E-07	3.64E-07	3.64E-07	3.70E-07	3.70E-07	3.11E-07	3.08E-07	4.12E-07	4.15E-07
Scenario 4 Pops Preferred Model Parameter	East Pamir Kyrgyz		Lowland Kyrgyz		Lowland Tajik		Pamiri Tajik		Sarikoli Tajik		Wakhi Tajik	
	Split-Admixture		Admixture-Admixture		Split-Admixture		Split-Admixture		Split-Admixture		Split-Admixture	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
NEast	6.23E+04	6.17E+04	7.07E+04	7.09E+04	3.59E+04	3.11E+04	7.12E+04	7.14E+04	6.80E+04	6.79E+04	4.13E+04	3.76E+04
NWest	8.13E+04	8.29E+04	7.23E+04	7.27E+04	7.20E+04	7.23E+04	8.74E+04	8.84E+04	7.71E+04	7.80E+04	6.00E+04	5.94E+04
NTarget	2.27E+04	2.10E+04	3.07E+04	3.12E+04	2.42E+04	2.22E+04	1.72E+04	1.48E+04	2.21E+04	2.03E+04	3.52E+04	3.61E+04
NCA	2.30E+04	2.19E+04	2.85E+04	2.96E+04	2.11E+04	1.85E+04	2.37E+04	2.30E+04	2.64E+04	2.66E+04	2.86E+04	2.99E+04
t ^{1a}	4.19E+02	3.36E+02	4.58E+02	3.76E+02	8.48E+02	8.65E+02	7.01E+02	6.42E+02	2.82E+02	2.33E+02	7.37E+02	6.74E+02
r1	3.53E-01	3.04E-01	6.65E-01	7.22E-01	3.53E-02	1.84E-02	1.76E-01	1.16E-01	1.97E-01	1.42E-01	1.64E-01	1.02E-01
t ^{2a}	7.09E+02	5.93E+02	1.03E+03	9.94E+02	8.14E+02	6.99E+02	1.19E+03	1.22E+03	6.76E+02	5.50E+02	1.41E+03	1.47E+03
r2	na	na	4.67E-01	4.43E-01	na	na	na	na	na	na	na	na
td ^a	3.03E+03	3.04E+03	3.01E+03	3.03E+03	3.21E+03	3.29E+03	2.95E+03	2.93E+03	2.92E+03	2.89E+03	2.84E+03	2.79E+03
NAncesor	5.07E+02	3.97E+02	5.51E+02	4.18E+02	2.65E+02	2.14E+02	7.40E+02	5.73E+02	5.77E+02	4.47E+02	4.73E+02	3.56E+02
μ ^b	3.39E-07	3.36E-07	3.47E-07	3.45E-07	3.63E-07	3.62E-07	3.23E-07	3.21E-07	3.43E-07	3.41E-07	4.40E-07	4.43E-07

Note: ^ain generations, ^bsubstitution per site per generation

suggesting the genetic differentiation is small. It is in agreement that the migrations of Kyrgyz to the Pamirs are recent [5]. More interestingly, we detect substantial genetic differentiation among three highland Tajik populations. The PCA shows Wakhi Tajik from the Wakhan Corridor, the southern edge of Pamirs, separating from Pamiri Tajik and Sarikoli Tajik (Fig. 3). Wakhi Tajik underwent a different demographic history in which no obvious female effective size increasing happened within 50 kya (Fig. 4). It should be noted that, the split time of Wakhi Tajik is before the LGM (Table 3), which is largely in disagreement with the dating of mtDNA lineages (Fig. 2 and Table S5). In addition to the setting of scenarios and models, the split time estimation relies on the selection of the representative West Eurasian. The lineages of haplogroup W3 account for 12.1% in Wakhi Tajik (Table S3), but are absent in Persian. So the coalescent time for haplogroup W lineages of Wakhi Tajik and Persian would be the age of haplogroup W, i.e., before the LGM [30]. As a result, the split time estimation is overestimated. And it cannot be simply explained as the time for the origin of Wakhi Tajik or the initial settlement of the Pamirs. As for Pamiri Tajik and Sarikoli Tajik from West and East Pamirs, respectively, the ABC analyses detect stronger gene flow from East Eurasian to Sarikoli Tajik than that to Pamiri Tajik and Wakhi Tajik (Table 3). Thus, the mtDNA heterogeneity in Pamir highlanders probably mirrors the complex history for peopling of Pamirs from the matrilineal genetic perspective.

Furthermore, Jeong et al. proposed that genetic admixture mediated selective signals for high-altitude adaptation in Tibetan Plateau [46]. Some studies showed positive selection on mtDNA lineages in Tibetans [47] and Sherpas [48]. Our results indicate that the Pamir highlanders are admixed populations. It gives us an opportunity to test whether admixture can mediate selection on mtDNAs in the Pamir highlanders or not. The comparisons of the N/S ratio between the highlanders and lowlanders/general population (Fig. S5 and Table S7) provide no evidence to supporting positive selection or relaxation of selective constraints in the highlanders. The patterns are different from those observed in Tibetans and Sherpas from Tibetan Plateau. It is expected that, future analyses based on extensive samplings (e.g., ref. [49]), genome-wide genetic markers (e.g., ref. [46]), even together with ancient DNAs (e.g., ref. [49]), will provide more details about the admixture history in the Pamirs.

Acknowledgments We are grateful to all volunteers and samplers. We thank Ni-Ni Shi, Quan-Kuan Shen, Ya-Jiang Wu, Mu-Yang Wang, Yao-Ming Li, and Feng Xu for the technical assistance. This study was supported by grants from National Natural Science Foundation of China (31301026), Science & Technology Department of Xinjiang Uygur Autonomous Region (201491188), Research Center for Ecology and Environment of Central Asia, Chinese Academy of

Sciences, and Bureau of Science and Technology of Yunnan Province. M.-S.P. thanks the support from the Youth Innovation Promotion Association, Chinese Academy of Sciences.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

1. Simonson TS. Altitude adaptation: a glimpse through various lenses. *High Alt Med Biol.* 2015;16:125–37.
2. Keay J. When men and mountains meet: the explorers of the Western Himalayas 1820–1875. 1st ed. Karachi: Oxford University Press; 1994.
3. Aldashev A, Naeije R, (eds). Problems of high altitude medicine and biology. Dordrecht: Springer; 2007.
4. Liu M. Migration and adaptation: the ethnography of the Tajik in the Pamirs. Beijing: Social Sciences Academic Press; 2014.
5. Dani AH, Masson VM. History of civilizations of Central Asia: development in contrast: from the sixteenth to the mid-nineteenth century. UNESCO; 2003.
6. Baumer C. The history of Central Asia: the age of the Steppe Warriors. I.B. Tauris; 2012.
7. Kuz'mina EE, Mallory JP. The origin of the Indo-Iranians. Brill; 2007.
8. Han K. The racial anthropological study of the ancient Xinjiang inhabitants. Urumqi: Xinjiang People's Publishing House; 1994.
9. Comas D, Calafell F, Mateu E, et al. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet.* 1998;63:1824–38.
10. Comas D, Plaza S, Wells RS, et al. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet.* 2004;12:495–504.
11. Heyer E, Balaesque P, Jobling MA, et al. Genetic diversity and the emergence of ethnic groups in Central Asia. *BMC Genet.* 2009;10:49.
12. Ricaut FX, Keyser-Tracqui C, Cammaert L, Crubézy E, Ludes B. Genetic analysis and ethnic affinities from two Scytho-Siberian skeletons. *Am J Phys Anthropol* 2004;123:351–60.
13. Li C, Li H, Cui Y, et al. Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biol* 2010;8:15.
14. Ovchinnikov IV, Malek MJ, Drees K, Kholina OI. Mitochondrial DNA variation in Tajiks living in Tajikistan. *Leg Med.* 2014;16:390–95.
15. Duggan AT, Stoneking M. A highly unstable recent mutation in human mtDNA. *Am J Hum Genet.* 2013;92:279–84.
16. Ning C, Gao S, Deng B, et al. Ancient mitochondrial genome reveals trace of prehistoric migration in the east Pamir by pastoralists. *J Hum Genet.* 2016;61:103–8.
17. Schönberg A, Theunert C, Li M, Stoneking M, Nasidze I. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur J Hum Genet.* 2011;19:988–94.
18. Derenko M, Malyarchuk B, Bahmanimehr A, et al. Complete mitochondrial DNA diversity in Iranians. *PLoS ONE.* 2013;8: e80673.
19. Fendt L, Zimmermann B, Daniaux M, Parson W. Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. *BMC Genomics.* 2009;10:139.
20. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge

- reference sequence for human mitochondrial DNA. *Nat Genet.* 1999;23:147–47.
21. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14:178–92.
 22. Weissensteiner H, Pacher D, Kloss-Brandstätter A, et al. Haplo-Grep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 2016;44:W58–63.
 23. Fan L, Yao YG. An update to MitoTool: using a new scoring system for faster mtDNA haplogroup determination. *Mitochondrion.* 2013;13:360–63.
 24. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 2009;30:E386–94.
 25. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10:564–67.
 26. Liu J, Wang LD, Sun YB, et al. Deciphering the signature of selective constraints on cancerous mitochondrial genome. *Mol Biol Evol.* 2012;29:1255–61.
 27. Forster P, Harding R, Torroni A, Bandelt HJ. Origin and evolution of native American mtDNA variation: a reappraisal. *Am J Hum Genet.* 1996;59:935–45.
 28. Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet.* 2000;67:718–26.
 29. Perego UA, Achilli A, Angerhofer N, et al. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol.* 2009;19:1–8.
 30. Soares P, Ermini L, Thomson N, et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet.* 2009;84:740–59.
 31. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 2012;29:1969–73.
 32. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012;9:772.
 33. Cornuet JM, Pudlo P, Veyssier J, et al. DIYABCv2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics.* 2014;30:1187–89.
 34. Seo SB, Zeng X, King JL, et al. Underlying data for sequencing the mitochondrial genome with the massively parallel sequencing platform Ion Torrent PGM. *BMC Genomics.* 2015;16:S4.
 35. Avise JC. *Phylogeography: the history and formation of species.* Cambridge: Harvard University Press; 2000.
 36. Atkinson QD, Gray RD, Drummond AJ. mtDNA variation predicts population size in humans and reveals a major southern Asian chapter in human prehistory. *Mol Biol Evol.* 2008;25:468–74.
 37. Clark PU, Dyke AS, Shakun JD, et al. The last glacial maximum. *Science.* 2009;325:710–14.
 38. Heller R, Chikhi L, Siegmund HR. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS ONE.* 2013;8:e62992.
 39. Glantz MM. The history of hominin occupation of central Asia in review. In: Norton CJ, Braun DR, editors. *Asian paleoanthropology: from Africa to China and beyond.* Dordrecht: Springer; 2010.
 40. Krause J, Orlando L, Serre D, et al. Neanderthals in Central Asia and Siberia. *Nature.* 2007;449:902–4.
 41. Sawyer S, Renaud G, Viola B, et al. Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc Natl Acad Sci USA.* 2015;112:15696–700.
 42. Ding Q, Hu Y, Xu S, Wang J, Jin L. Neanderthal introgression at chromosome 3p21.31 was under positive natural selection in East Asians. *Mol Biol Evol.* 2014;31:683–95.
 43. Currat M, Excoffier L. Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. *Proc Natl Acad Sci USA.* 2011;108:15129–34.
 44. Wells RS, Yuldasheva N, Ruzibakiev R, et al. The Eurasian Heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA.* 2001;98:10244–49.
 45. Fregel R, Cabrera V, Larruga JM, Abu-Amero KK, González AM. Carriers of mitochondrial DNA macrohaplogroup N lineages reached Australia around 50,000 years ago following a northern Asian route. *PLoS ONE.* 2015;10:e0129839.
 46. Jeong C, Alkorta-Aranburu G, Basnyat B, et al. Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat Commun.* 2014;5:3281.
 47. Gu M, Dong X, Shi L, Lin K, Huang X, Chu J. Differences in mtDNA whole sequence between Tibetan and Han populations suggesting adaptive selection to high altitude. *Gene.* 2012;496:37–44.
 48. Kang L, Zheng HX, Chen F, et al. mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. *Mol Biol Evol.* 2013;30:2579–87.
 49. Olivieri A, Sidore C, Achilli A, et al. Mitogenome diversity in Sardinians: a genetic window onto an island's past. *Mol Biol Evol.* 2017;34:1230–39.
 50. Hey J. On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biol.* 2005;3:e193.

Affiliations

Min-Sheng Peng¹ · Weifang Xu^{2,3} · Jiao-Jiao Song^{1,4} · Xing Chen¹ · Xierzhatijiang Sulaiman⁵ · Liuhong Cai⁶ · He-Qun Liu¹ · Shi-Fang Wu¹ · Yun Gao¹ · Najmudinov Tojiddin Abdulloevich⁷ · Manilova Elena Afanasevna⁷ · Khudoidodov Behruz Ibrohimovich⁷ · Xi Chen^{8,9} · Wei-Kang Yang^{8,9} · Miao Wu^{8,9} · Gui-Mei Li¹⁰ · Xing-Yan Yang¹¹ · Allah Rakha^{12,13} · Yong-Gang Yao^{12,14,15} · Halmurat Upur² · Ya-Ping Zhang^{1,11,14,15}

¹ State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

² Key Laboratory of the Chinese Ministry of Education and Xinjiang Uighur Autonomous Region for High-Incident Diseases in Uighur Ethnic Population, Xinjiang Medical University, Urumqi 830011, China

³ Traditional Chinese Medicine Hospital Affiliated to Xinjiang Medical University, Urumqi 830000, China

⁴ Institute of Health Sciences, Anhui University, Hefei 230601, China

⁵ College of Uighur Medicine, Xinjiang Medical University, Urumqi 830011, China

- ⁶ The Second People's Hospital of Kashi, Kashi 844000, China
- ⁷ E.N. Pavlovsky Institute of Zoology and Parasitology, Academy of Sciences of Republic of Tajikistan, Dushanbe 734025, Tajikistan
- ⁸ Research Center for Ecology and Environment of Central Asia, Chinese Academy of Sciences, Urumqi 830011, China
- ⁹ Key Laboratory of Biogeography and Bioresource in Arid Land, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China
- ¹⁰ Kunming Biological Diversity Regional Center of Large Apparatus and Equipments, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
- ¹¹ State Key Laboratory for Conservation and Utilization of Bio-Resources, Yunnan University, Kunming 650091, China
- ¹² Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming 650223, China
- ¹³ Department of Forensic Sciences, University of Health Sciences, Lahore 54600, Pakistan
- ¹⁴ KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming 650223, China
- ¹⁵ Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China