



HAL
open science

In the heartland of Eurasia: the multilocus genetic landscape of Central Asian populations

Evelyne Heyer, Begoña Martínez-Cruz, Renaud Vitalis, Laure Ségurel, Frédéric Austerlitz, Myriam Georges, Sylvain Théry, Lluís Quintana-Murci, Tatyana Hegay, Almaz Aldashev, et al.

► To cite this version:

Evelyne Heyer, Begoña Martínez-Cruz, Renaud Vitalis, Laure Ségurel, Frédéric Austerlitz, et al.. In the heartland of Eurasia: the multilocus genetic landscape of Central Asian populations. *European Journal of Human Genetics*, 2010, 10.1038/ejhg.2010.153 . hal-00574375

HAL Id: hal-00574375

<https://hal.science/hal-00574375>

Submitted on 8 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

25 ¹⁰These authors contributed equally to the present study

26

27 *Correspondence:

28 Evelyne Heyer

29 e-mail: heyere@mnhn.fr

30 Phone: +33 (0)1 40 79 81 58

31 Fax: +33 (0)1 40 79 32 31

32

33 Running title: Multilocus genetic landscape in Central Asia

34 Keywords: admixture; Central Asia; ethnic groups; genetic diversity; microsatellites;

35 population genetics

36

37 **Abstract**

38 Located in the Eurasian heartland, Central Asia has played a major role in both the early
39 spread of modern humans out of Africa and the more recent settlements of differentiated
40 populations across Eurasia. A detailed knowledge of the peopling in this vast region would
41 therefore greatly improve our understanding of range expansions, colonizations, and recurrent
42 migrations, including the impact of the historical expansion of eastern nomadic groups that
43 occurred in Central Asia. However, despite its presumable importance, little is known about
44 the level and the distribution of genetic variation in this region. We genotyped 26 Indo-
45 Iranian- and Turkic-speaking populations, belonging to six different ethnic groups, at 27
46 autosomal microsatellite loci. The analysis of genetic variation reveals that Central Asian
47 diversity is mainly shaped by linguistic affiliation, with Turkic-speaking populations forming
48 a cluster more closely related to East Asian populations and Indo-Iranian speakers forming a
49 cluster closer to Western Eurasians. The scattered position of Uzbeks across Turkic- and
50 Indo-Iranian speaking populations may reflect their origins from the union of different tribes.
51 We propose that the complex genetic landscape of Central Asian populations results from the
52 movements of eastern, Turkic-speaking groups during historical times, into a long lasting
53 group of settled populations, which may be represented nowadays by Tajiks and Turkmen.
54 Contrary to what is generally thought, our results suggest that the recurrent expansions of
55 eastern nomadic groups did not result in the complete replacement of local populations but
56 rather into partial admixture.

57

58 **Introduction**

59 The evolutionary history of modern humans has been characterized by range expansions,
60 colonizations and recurrent migrations over the last 100,000 years.¹ Some regions of the
61 world that have served as natural corridors between landmasses are of particular importance
62 in the history of human migrations. Central Asia is probably at the crossroads of such
63 migration routes.^{1,2} Located in the Eurasian heartland, it encompasses a vast territory, limited
64 to the east by the Pamir and Tien-Shan mountains, to the west by the Caspian Sea, to the north
65 by the Russian taiga and to the south by the Iranian deserts and Afghan mountains. The role
66 of Central Asia in both the early spread of modern humans out of Africa and the more recent
67 settlement of differentiated populations³ is not precisely known.⁴⁻⁶ For example, it remains
68 unclear whether this region harbored a Palaeolithic "maturation phase" of modern humans
69 before giving rise to waves of migration resulting in colonization of the Eurasian continent⁶ or
70 whether it has served as a meeting place for previously differentiated Asian and European
71 populations following their initial expansions.^{3,7}

72 Central Asia entered the historical records about 1300 B.C., when Aryan tribes
73 invaded the Iranian territory from what is nowadays Turkmenistan and established the Persian
74 Empire in the 7th Century B.C.⁸ A branch of those, the Scythians, described in ancient
75 Chinese texts and in Herodotus' *Histories* as having European morphological traits and
76 speaking Indo-Iranian languages, expanded north into the steppes. Thereafter, Central Asia
77 was faced with multiple waves of Turkic migrations, although it is difficult to know precisely
78 when these westward expansions began. Between the second and the first century B.C., Huns
79 brought the East-Asian anthropological phenotype to Central Asia.⁸ At the same period, the
80 Chinese established a trade route (the Silk Road), which connected the Mediterranean Basin
81 and Eastern Asia for more than 16 centuries. In the 13th century A.D. the Turco-Mongol
82 Empire lead by Genghis Khan became the largest of all time, from Mongolia to the Black Sea.

83 All these movements of populations resulted in a considerable ethnic diversity in Central
84 Asia, with Indo-Iranian speakers living as sedentary agriculturalists and Turkic speakers
85 mainly living as traditionally nomadic herders.

86 Together with the ancient peopling of Central Asia, this intricate demographic history
87 shaped patterns of genetic variability in a complex manner. Most previous studies, based on
88 classical markers,¹ mitochondrial DNA (mtDNA)^{3,9-13} or the non-recombining portion of the
89 Y-chromosome (NRY),^{6,14-16} have shown that genetic diversity in Central Asia is among the
90 highest in Eurasia.^{3,6,15} NRY studies suggest an early settlement of Central Asia by modern
91 humans, followed by subsequent colonization waves in Eurasia,⁶ while some mtDNA studies
92 point to an admixed origin from previously differentiated Eastern and Western Eurasian
93 populations.¹¹ Furthermore, a recent analysis of mtDNA data suggests east-to-west
94 expansions waves across Eurasia.¹⁴ However, inferring more accurately the impact of
95 population movements, including the expansion of eastern nomadic groups, requires
96 additional, fast-evolving molecular markers. Here we report on the first multilocus autosomal
97 genetic survey of Central Asian populations. Twenty-six populations from six ethnic groups
98 were genotyped at 27 autosomal unlinked microsatellite markers. We aimed to shed light on
99 the genetic origins of Central Asian populations, and to investigate how the recurrent
100 westward expansions of eastern nomadic groups during historical times have shaped the
101 Central Asian genetic landscape.

102

103 **Materials and methods**

104 **DNA samples**

105 We sampled 767 men belonging to 26 populations from western Uzbekistan to eastern
106 Kyrgyzstan (Table 1 and Figure 1) representative of the ethnological diversity in Central
107 Asia: Tajiks, which are Indo-Iranian speakers (a branch of the Indo-European language
108 family) and Kazakhs, Turkmen, Karakalpaks, Kyrgyz and Uzbeks, which are Turkic speakers
109 (a branch of the Altaic language family). In two Uzbek populations from the Bukhara area
110 (LUZa and LUZn), an extensive linguistic survey showed that individuals were bilingual,
111 speaking both Tajik and Uzbek. Since their home language was Tajik (an Indo-Iranian
112 language), we further classified these two populations into the Indo-Iranian group for
113 subsequent analyses. We collected individuals unrelated for at least two generations back in
114 time. All individuals gave informed consent for their participation in this study. Total
115 genomic DNA was isolated from blood samples by a standard salting out procedure¹⁷
116 followed by a phenol-chloroform extraction.¹⁸

117

118 **Genotyping**

119 We selected 27 microsatellite markers¹⁹ from the set of 377 markers used in the worldwide
120 study by Rosenberg *et al.*²⁰ The choice and description of markers, PCR and electrophoresis
121 conditions are given in Ségurel *et al.*¹⁹ We further genotyped 20 individuals from the HGDP-
122 CEPH Human Genome Diversity Cell Line Panel²⁰⁻²² at the 27 microsatellite loci, in order to
123 standardize the original Central Asian data presented here with the worldwide HGDP-CEPH
124 data.

125

126 **Data analyses**

127 Genetic diversity

128 In each population and for each locus, we calculated the allelic richness (AR) using the
129 rarefaction method proposed by El Mousadik *et al.*²³ with the software package FSTAT.²⁴
130 Unbiased estimates of expected heterozygosity (H_e)²⁵ were computed in each population for
131 each locus with GENETIX.²⁶ Both AR and H_e estimates were averaged over loci in each
132 population. We tested heterogeneity in both AR and H_e among populations using the Kruskal-
133 Wallis test, with locus-specific estimates taken as replicate observations. Locus-specific
134 allelic richness and expected heterozygosity were also estimated for populations pooled into
135 Indo-Iranian- and Turkic-speaking groups, and averaged over loci within groups. We tested
136 between-group differences in both AR and H_e using the Wilcoxon's signed-rank test, with
137 locus-specific estimates taken as replicate observations. We further estimated AR and H_e for
138 each locus over the pooled data from Central Asia and over the pooled data for Central/South
139 Asia, East Asia, Europe and the Middle-East from the HGDP-CEPH Panel, and calculated the
140 averages over loci within groups. We tested heterogeneity in both AR and H_e across the five
141 groups of Eurasian populations using the Kruskal-Wallis test, taking locus-specific estimates
142 as replicate observations. When significant differences among groups were found, we ran the
143 Tukey range test to find which group statistics were significantly different from one another.
144 All statistical analyses were performed with the software package JMP5.1 (SAS Institute
145 Inc.).²⁷

146

147 Genetic structure

148 Population differentiation (F_{ST}) was calculated overall and between pairs of Central Asian
149 populations with GENEPOP 4.0.²⁸ Exact tests of differentiation were performed with
150 FSTAT,²⁴ adjusting p -values with Bonferroni correction for multiple tests. We performed a
151 correspondence analysis (CA) based on tables of allele counts using GENETIX.²⁶ The
152 population structure was also inferred by means of a hierarchical analysis of molecular

153 variance (AMOVA²⁹), with populations pooled into ethnic or linguistic groups. For ethnic
154 grouping, populations were pooled as Tajiks (TJA, TDS, TJT, TJK, TJR, TJN, TDU, TJE,
155 TJY and TJU), Karakalpaks (KKK and OTU), Kazakhs (KAZ and LKZ), Kyrgyz (KRA,
156 KRG, KRL, KRB, KRT and KRM), Uzbeks (UZA, UZB, LUZa, LUZn and UZT) and
157 Turkmen (TUR). For linguistic grouping, populations were pooled as Indo-Iranian speakers
158 (Tajiks and the two Uzbek populations LUZa and LUZn) and Turkic speakers (all other
159 populations). These analyses were performed with ARLEQUIN 3.11.³⁰ Isolation-by-distance
160 (IBD) was tested with GENEPOP 4.0.²⁸ We used PATHMATRIX³¹ to compute the matrix of
161 effective geographical distances, based on a least-cost path algorithm. The least-cost
162 distances, which account for the cost of the movement through the slopes in the landscape,
163 were calculated from the digital elevation model GTOPO30 of the Earth Resources
164 Observation and Science (EROS) Center.

165

166 Clustering analyses

167 We performed a clustering analysis with STRUCTURE³² on the Central Asian populations
168 together with all the Eurasian and African populations from the HGDP-CEPH Panel H952
169 corrected dataset.^{33,34} We used the latest version of STRUCTURE³⁵ (version 2.3), which
170 allows structure to be detected at lower levels of divergence than the original model. Each
171 Markov chain was run for 10^6 steps, after a 10^5 -step burn-in period. In each case, the results
172 were checked to ensure consistency over forty independent runs. Potential distinct modes
173 among the 40 runs were identified using the *Greedy* algorithm implemented in CLUMP³⁶. We
174 varied the hypothetical number of clusters (K) from 1 to 8 for all analyses. All chains were
175 run using the F model for correlations of allele frequencies across clusters.³⁷

176

177 Admixture analyses

178 The Central Asian genetic pool may be more than just the result of admixture from Eurasian
179 populations, but we were nonetheless interested in investigating the potential origins of
180 Central Asian populations among all Eurasian populations. We used LEADMIX³⁸ to calculate
181 maximum likelihood estimates (MLE) of the admixture proportions for each Central Asian
182 population. We ran the program independently for each of them, considering four putative
183 parental groups from the HGDP-CEPH Panel: Central/South Asia, East Asia, Europe and
184 Middle East. For the Central/South Asian group, we chose a pool of Balochi ($n = 25$) and
185 Makrani ($n = 25$) individuals, both populations being non-significantly differentiated ($F_{ST} = -$
186 0.002 ; exact test $p = 0.34$). We chose the Han Chinese ($n = 44$) for the East Asian parental
187 group, and we further considered a pool of French ($n = 28$), Bergamo ($n = 13$) and Tuscan (n
188 $= 21$) individuals for the European group, these three populations being non-significantly
189 differentiated ($F_{ST} < -0.006$; $p > 0.42$). Last, we chose the Palestinians ($n = 46$) for the Middle
190 Eastern group.³⁹

191

192 **Results**

193 **Genetic diversity**

194 Average allelic richness and expected heterozygosity for each of the 26 Central Asian
195 populations and across regions are given in Table 2. We found a significant difference in
196 allelic richness (Kruskal-Wallis test, $\chi^2 = 105,29$, d.f. = 25, $p < 0.0001$) and in expected
197 heterozygosity (Kruskal-Wallis test, $\chi^2 = 67.98$, d.f. = 25, $p < 0.0001$) among populations. We
198 found no significant difference in allelic richness between Indo-Iranian ($AR = 13.8$) and
199 Turkic speakers ($AR = 13.7$, Wilcoxon signed rank test, $Z = -0.69$, $p = 0.49$), although the
200 expected heterozygosity was significantly higher in Indo-Iranian as compared to Turkic
201 speakers ($H_e = 0.818$ and $H_e = 0.787$, respectively, Wilcoxon signed rank test, $Z = -4.55$, $p <$
202 0.0001). We found a significant difference in allelic richness across Central Asia, Europe,
203 Central/South Asia, Middle East and East Asia (Kruskal-Wallis test, $K = 36.46$, d.f. = 4, $p <$
204 0.0001), as well as in expected heterozygosity (Kruskal-Wallis test, $K = 52.94$, d.f. = 4, $p <$
205 0.0001). Yet, these differences were rather due to a lower heterozygosity in East Asia and also
206 slightly higher allelic richness in Middle East (Tukey's test, $p < 0.0001$ for both AR and H_e).
207 Central Asia therefore showed neither higher nor lower diversity than the rest of Eurasia.

208

209 **Population differentiation**

210 The 26 Central Asian populations were slightly but significantly differentiated ($F_{ST} = 0.015$,
211 $CI_{99\%} = [0.011-0.018]$, $p < 0.01$). Pairwise F_{ST} estimates ranged from -0.004 to 0.056, with
212 205 out of 325 pairs of populations (i.e., 63.1%) being significantly differentiated after
213 Bonferroni correction for multiple tests (see Supplementary Table 1). These significant
214 estimates mainly corresponded to pairwise comparisons between one Turkic and one Indo-
215 Iranian population, as well as to comparisons between two Indo-Iranian populations. The
216 apportionment of genetic variation among linguistic or ethnic groups of populations (Table 3)

217 showed that more than 98% of the total variation lay within populations ($p < 0.0001$). Yet,
218 both ethnicity and linguistic affiliation accounted significantly for the observed variation (F_{CT}
219 = 0.007, $p < 0.0001$ and $F_{CT} = 0.011$, $p < 0.0001$, respectively). We found no evidence of
220 isolation-by-distance within each of Turkic and Indo-Iranian group of populations ($p = 0.363$
221 and $p = 0.772$, respectively).

222 The correspondence analysis (CA) based on the table of allele counts in Central Asia
223 separated Turkic- and Indo-Iranian-speaking populations on the first axis (Figure 2a). The
224 first two factorial components (FC) accounted for 20.5 % of the total inertia. There were some
225 exceptions, though: two Turkic-speaking populations, TUR and UZA, were clearly clustered
226 with Indo-Iranian-speaking populations. Interestingly, the Uzbek populations (LUZa, LUZn,
227 UZA and UZT) showed a scattered pattern on the CA which overlapped the Turkic-speaking
228 and the Indo-Iranian-speaking groups of populations. The CA based on the table of allele
229 counts in Eurasia placed Central Asian populations in an intermediate position between a
230 group of European population, a group of Middle Eastern populations, a group of
231 Central/South Asian populations, and a group of East Asian populations (Figure 2b). The first
232 two factorial components accounted for 22.4 % of the total inertia. Turkic- and Indo-Iranian-
233 speaking populations were separated on the first axis, with Turkic-speaking populations being
234 closer to East Asian populations, and Indo-Iranian-speaking populations being closer to
235 Central/South Asian, European and Middle Eastern populations. It is noteworthy that Central
236 Asian and Central/South Asian populations were more scattered than any other group of
237 populations in Eurasia (Figure 2b). Interestingly, the Hazaras from Pakistan, who claim to be
238 direct male-line descendants of Genghis Khan,^{40,41} as well as the Uygurs, clustered together
239 with the Turkic-speaking populations of Central Asia.

240

241 **Cluster analyses**

242 Analyzing the Eurasian plus the African populations altogether, we found that the highest
243 average posterior probability of the data (D), across 40 runs, was obtained for $K = 7$ putative
244 clusters, with $\text{Log}[P(K = 7 | D)] = -167565.4$ (SD = 22.8), although the average posterior
245 probability for $K = 6$ was only slightly lower, with $\text{Log}[P(K = 6 | D)] = -167653.8$ (SD =
246 10.6). The symmetric similarity coefficients computed with CLUMPP across independent
247 runs were all larger than 0.99 for K varying from 2 to 5, and larger than 0.87 for $K = 6$, which
248 suggests the absence of genuine multimodality across runs. As seen in Figure 3, at $K = 2$, we
249 observed a clear east-west cline. Central Asia seemed to be intermediate between one cluster
250 made of European, Middle Eastern, Central/South Asian and African populations on the one
251 hand and one cluster of East Asian populations on the other hand, which is consistent with the
252 CA (Figure 2b). There was no individual assigned exclusively to one cluster, with Turkic-
253 speaking individuals having a higher membership coefficient in the East Asian cluster, and
254 Indo-Iranian-speaking individuals having a higher membership coefficient in the cluster made
255 of Europe, Middle East, Central/South Asia and Africa. At $K = 3$, the six African populations
256 clustered together. At $K = 4$, the European and Middle Eastern populations clustered together,
257 with Central/South Asian and Central Asian populations (mostly Indo-Iranian speakers)
258 showing a small contribution from this European/Middle Eastern cluster (represented in green
259 in Figure 3). At $K = 5$, the Turkic-speaking populations from Central Asia showed a large
260 contribution from a fifth cluster (in orange in Figure 3). At $K = 6$, the Indo-Iranian speaking
261 populations from Central Asia show a large contribution from a sixth cluster (in light blue in
262 Figure 3). The two latter clusters were found almost exclusively in Central Asian populations.
263 Most Turkic-speaking populations showed a contribution from the East Asian cluster (in red),
264 and most Indo-Iranian populations showed a contribution from Europe and Middle East (in
265 green). It is noteworthy that Uygur and Hazara populations showed the same pattern as the
266 Turkic-speaking populations from Central Asia. At $K = 7$, all Eurasian populations (but

267 mostly Turkic-speaking populations) had a variable proportion of the new component. Yet, no
268 run at $K = 7$ resulted in a new cluster of populations, as compared to $K = 6$, which is the
269 reason why the output for $K = 7$ is not represented in Figure 3.

270

271 **Admixture analyses**

272 The maximum likelihood estimates (MLE) of admixture proportions obtained with
273 LEADMIX for each Central Asian population are given in Figure 1 and Table 4. Most Turkic-
274 speaking populations had a large East Asian ancestral contribution, which represented in
275 general 49.5%, or more, of the total contribution. There were four notable exceptions, though,
276 with the Turkmen (TUR) and three Uzbek populations (UZA, UZB and UZT) showing a
277 lower contribution from East Asian populations (respectively, 27.2%, 28.6%, 28.1% and
278 28.7%). Indo-Iranian-speaking populations had a large western Eurasian contribution
279 (Central/South Asia, Europe and Middle-East), which represented 72.7% to 94.5% of the total
280 contribution, although the relative contributions from these three parental groups differ across
281 Indo-Iranian-speaking populations. It is noteworthy that, in general, many geographically
282 close populations that speak different languages showed contrasted admixture proportions
283 (see, e.g., UZT and TJU in Table 4), which supports the idea that language is a major
284 determinant of population differentiation in Central Asia.

285

286 **Discussion**

287 **Central Asia in the heartland of Eurasia**

288 We found a high level of autosomal genetic diversity in Central Asia, consistent with previous
289 observations,^{3,16} and similar in extent to other major regions in Eurasia (Table 2). Population
290 differentiation among Central Asian populations was similar, or even stronger, than that
291 measured among populations within other regions in Eurasia: the pairwise F_{ST} estimates
292 ranged from -0.004 to 0.056 in Central Asia, a range which should be compared to that found
293 in the European group [-0.011; 0.015], the Middle-Eastern group [0.008; 0.021], the
294 Central/South Asian group [-0.002; 0.062] and in the East Asian group [-0.011; 0.046], based
295 on the same set of 27 microsatellite loci as we used in our study. This pattern is also apparent
296 in the correspondence analysis (Figure 2b), where Central Asian and Central/South Asian
297 populations were more scattered than each of the East Asian, European and Middle-Eastern
298 groups, which suggests a higher diversification within Central Asia and Central/South Asia.
299 Most importantly, the observed diversity was mainly due to the differentiation into two main
300 groups of populations (Figure 3): on the one hand, Indo-Iranian-speaking populations (which
301 include Tajiks and three Uzbek populations) that are genetically closer to populations from
302 Western Eurasia; on the other hand, Turkic-speaking populations (which include Karakalpaks,
303 Kazakhs, Kyrgyz, and two other Uzbek populations) that are closer to Eastern Asian
304 populations (with the exception of the Turkmen). This pattern was also apparent in the
305 correspondence analysis (Figure 2b), and consistent with the significant differentiation of
306 almost all pairwise comparisons between an Indo-Iranian-speaking and a Turkic-speaking
307 population (Supplementary Table 1).

308 Although several studies have shown that geography is, in general, a better predictor
309 of genetic differentiation than ethnicity and linguistics,^{42,43} language affiliation appears as the
310 most important factor explaining the distribution of genetic diversity in Central Asia (Table

311 3). We found indeed that, although most (98%) of the variation lay within Central Asian
312 populations ($p < 0.0001$), a significant part of the total variation (1.09%; $p < 0.0001$) lay
313 among linguistic groups, which provides an estimate of differentiation among groups equal to
314 $F_{CT} = 0.011$. For comparison purpose, the differentiation among Central/South Asia, East
315 Asia, Europe and Middle East was found to be $F_{CT} = 0.044$, with 94.1% of the total variation
316 found within populations ($p < 0.0001$) and 4.4% found among groups, based on the same set
317 of 27 microsatellite loci as we used in our study. We found no evidence of a correlation
318 between geography and genetics within each of the Indo-Iranian or Turkic groups of Central
319 Asian populations. For the Turkic-speaking populations, this may be explained by their recent
320 arrival in the region and/or their nomadic life-style. However, more striking is the fact that no
321 geographic pattern of genetic variation was found among sedentary Indo-Iranian speakers
322 either.

323

324 **Putative origins of Indo-Iranian- and Turkic-speaking populations**

325 The clustering analysis showed that most individuals from the Indo-Iranian-speaking
326 populations had large membership coefficients into two clusters (light blue and beige in
327 Figure 3) that were found mostly in these populations. Altogether, the significant pairwise F_{ST}
328 estimates between almost all pairs of Indo-Iranian-speaking populations (Supplementary
329 Table 1), the high level of diversity across Indo-Iranian populations (Table 2) and the variable
330 level of admixture from the putative parental populations (Table 4) seem consistent with the
331 premise that Indo-Iranian speakers are long term settled populations in the area. This latter
332 hypothesis is strongly supported by archaeological evidence.⁴⁴ Conversely, we found a lower
333 genetic differentiation among Turkic-speaking populations despite their wide geographic
334 distribution (Figure 1), which suggests a more recent common origin of these populations as
335 compared to Indo-Iranian-speaking populations, in consistence with historical records.

336 Our study further shed some light on the origins of the Turkic-speaking populations in
337 Central Asia. The clustering analyses showed indeed that most individuals from the Turkic-
338 speaking populations had large membership coefficients into one Central Asian cluster (in
339 orange in Figure 3) and smaller membership coefficients into the East-Asian cluster (in red in
340 Figure 3) thus confirming the result of Li *et al*⁴⁵ based on a small central Asian cluster for
341 Uygur, Kazakh and Khanty. This pattern likely reflects the existence of an ancestral group of
342 Turkic-speakers (orange cluster in Figure 3), which popular Turkic culture considers as
343 originating from the Altai region. The East-Asian ancestry of Turkic-speaking populations
344 (red cluster in Figure 3) may then correspond to the westward expansions of nomadic groups
345 form East Asia during historical times.

346 The Westernized view of westward invasions usually emphasizes the extreme violence
347 and cruelty of the hordes led by Attila the Hun (A.D. 406-453), or that from the Mongolian
348 empire led by Genghis Khan. However, our results somehow challenge this view and rather
349 suggest that these more recent expansions did not lead to the massacre and complete
350 replacement of the locally settled populations but rather to partial admixture. We found
351 almost no eastern ancestry in Indo-Iranian speaking populations (see Figure 3), which
352 suggests that the group of people from which the current-day Tajik and Turkmen populations
353 would be the descendants, did not suffer from the westward expansions of eastern nomadic
354 groups. This is consistent with Zerjal *et al*'s study¹⁶, which showed the absence of the
355 “Genghis Khan lineage” in the Tajik and Turkmen populations they studied. Furthermore, the
356 present finding that the partial admixture with eastern nomadic groups concerned almost
357 exclusively the Turkic-speaking populations is consistent with the fact that Turks and
358 Mongols share cultural traditions and life-style, which may have facilitated inter-groups
359 marriages.

360 Our study also contradicts the claim that these westward invasions resulted in founder
361 effects.¹⁶ The high level of autosomal diversity observed in all Turkic-speaking populations
362 (Table 2) contrasts indeed with the low level of Y-chromosome diversity found in some
363 populations of the region.^{10,16} Our recent studies based on the analysis of uni-parental markers
364 in Central Asia already showed that the low level Y-chromosome diversity is only found in
365 the Turkic-speaking group⁴⁶, which may therefore be explained by the social organization of
366 Turkic-speaking populations, that is based on patrilineal descent groups.^{10, 18}

367 Overall, our results are partly consistent with Comas et al.'s hypothesis¹¹ that Central
368 Asia has been a contact zone between two differentiated groups. Our study suggests that one
369 of these groups is a long lasting group of settled populations, now represented by Tajiks and
370 Turkmen, although the origin of this group is difficult to infer; the second of these groups is
371 likely to have a more recent origin, resulting from the movements of eastern nomadic Turkic-
372 speaking groups. Interestingly, we found almost no African ancestry in the genetic pool of
373 Central Asian population from clustering analyses (Figure 3). Yet, with the same level of
374 clustering, we found no African ancestry either in Europe or in East-Asia. Further work is
375 therefore required to infer the more ancient peopling of Central Asia, after the spread of
376 modern humans out of Africa.

377 We found that the Uzbek populations were scattered across Turkic- and Indo-Iranian
378 speaking populations (Figure 2b). Some Uzbek populations (LUZa, LUZn, UZA) were closer
379 to Indo-Iranian speaking populations, while other populations (UZB, UZT) clearly clustered
380 with Turkic-speaking populations. This is consistent with the fact that Uzbek populations
381 include the 17th century Uzbeks, which were nomadic herders before they sedentarized
382 around the 16th Century,¹⁰ and the former Chagatai Turk groups who were already settled in
383 Uzbekistan.⁴⁷ Uzbeks therefore result from the union of different tribes, some of recent origin

384 clustering with Turkic-speaking populations, and some tracing back to Chagatai Turks who
385 were strongly admixed with Iranian dwellers of Central Asia.

386

387 **Evidence for linguistic replacements**

388 We found two presumable cases of linguistic replacements in Central Asia. The Turkic-
389 speaking populations, TUR (Turkmen) and UZA (Uzbek) were found to cluster together with
390 Indo-Iranian-speaking populations (Figure 2). The Uzbek population UZA, a currently
391 Turkic-speaking population, is indeed genetically more similar to Indo-Iranian speakers,
392 which suggests a linguistic shift in this population. Concerning the Turkmen, their genetic
393 similarity with Tajiks (see also Table 4) is consistent with the hypothesis that they may be the
394 present-day descendants of populations established over long periods of time. The indigenous
395 cultural history of the Turkmen in Turkmenistan can indeed be dated back to 10,000 years
396 B.C. and similarities between the cultures and technologies found in the archaeological record
397 suggest that this region has been continually occupied since 6,000 B.C. A recent linguistic
398 replacement in the TUR population would then explain the observed pattern of a Turkic-
399 speaking population clustering with Indo-Iranian speakers.

400

401 **A Central Asian origin of the Hazaras?**

402 Our study confirms the results of Li *et al.*'s study⁴⁸ that cluster the Hazara population with
403 Central Asian populations, rather than Mongolian populations, which is consistent with
404 ethnological studies.⁴⁹ Our results further extend these findings, since we show that the
405 Hazaras are closer to Turkic-speaking populations from Central-Asia, than to East-Asian or
406 Indo-Iranian populations.

407

408

409 **Acknowledgements**

410 We are indebted to everyone who volunteered to participate to this study. We also thank R.
411 Leblois and P. Verdu for insightful discussions on previous versions of this paper, H. Cann
412 for providing CEPH samples, the *Service de Systématique Moléculaire* (SSM) at the *Museum*
413 *national d'Histoire naturelle* (MNHN) for making facilities available, and J.A. Godoy for
414 technical assistance. We are very grateful to CESGA (Supercomputational Centre of Galicia)
415 and to the Computational Biology Service Unit from the *Museum national d'Histoire*
416 *naturelle* (MNHN – CNRS UMS 2700) where the computational analyses were performed.
417 This work was supported by the *Centre National de la Recherche Scientifique* (CNRS) ATIP
418 program (to E.H.), by the CNRS interdisciplinary program "*Origines de l'Homme du Langage*
419 *et des Langues*" (OHLL), the European Science Foundation (ESF) EUROCORES program
420 "The Origin of Man, Language and Languages" (OMLL) and the ANR grant
421 "NUTGENEVOL" (07-BLAN-0064).

422

423 **References**

- 424 1 Cavalli-Sforza LL, Menozzi P, Piazza A: The History and Geography of Human
425 Genes. Princeton, University Press, 1994.
- 426 2 Nei M, Roychoudhury AK: Evolutionary relationships of human populations on a
427 global scale. *Molecular Biology and Evolution* 1993; **10**: 927-943.
- 428 3 Comas D, Calafell F, Mateu E *et al*: Trading genes along the silk road: mtDNA
429 sequences and the origin of central Asian populations. *American Journal of Human*
430 *Genetics* 1998; **63**: 1824-1838.
- 431 4 Cordaux R, Deepa E, Vishwanathan H, Stoneking M: Genetic evidence for the demic
432 diffusion of agriculture to India. *Science* 2004; **304**: 1125-1125.
- 433 5 Karafet T, Xu LP, Du RF *et al*: Paternal population history of east Asia: Sources,
434 patterns, and microevolutionary processes. *American Journal of Human Genetics*
435 2001; **69**: 615-628.
- 436 6 Wells RS, Yuldasheva N, Ruzibakiev R *et al*: The Eurasian Heartland: A continental
437 perspective on Y-chromosome diversity. *Proceedings of the National Academy of*
438 *Sciences of the United States of America* 2001; **98**: 10244-10249.
- 439 7 Bowles G. The peoples of Asia; in: Nicolson Wa (ed). London, 1977.
- 440 8 Гумилев ЛНДГАСИ-ТНА-МН, 1967. - 504 с.. с карт. - 4800. 1967.
- 441 9 Chaix R, Austerlitz F, Khegay T *et al*: The genetic or mythical ancestry of descent
442 groups: Lessons from the Y chromosome. *American Journal of Human Genetics* 2004;
443 **75**: 1113-1116.
- 444 10 Chaix R, Quintana-Murci L, Hegay T *et al*: From social to genetic structures in central
445 Asia. *Current Biology* 2007; **17**: 43-48.

- 446 11 Comas D, Plaza S, Wells RS *et al*: Admixture, migrations, and dispersals in Central
447 Asia: evidence from maternal DNA lineages. *European Journal of Human Genetics*
448 2004; **12**: 495-504.
- 449 12 Lalueza-Fox C, Sampietro ML, Gilbert MTP *et al*: Unravelling migrations in the
450 steppe: mitochondrial DNA sequences from ancient Central Asians. *Proceedings of*
451 *the Royal Society of London Series B-Biological Sciences* 2004; **271**: 941-947.
- 452 13 Perez-Lezaun A, Calafell F, Comas D *et al*: Sex-specific migration patterns in central
453 Asian populations, revealed by analysis of Y-chromosome short tandem repeats and
454 mtDNA. *American Journal of Human Genetics* 1999; **65**: 208-219.
- 455 14 Chaix R, Austerlitz F, Hegay T, Quintana-Murci L, Heyer E: Genetic traces of east-to-
456 west human expansion waves in Eurasia. *American Journal of Physical Anthropology*
457 2008; **136**: 309-317.
- 458 15 Hammer MF, Karafet TM, Redd AJ *et al*: Hierarchical patterns of global human Y-
459 chromosome diversity. *Molecular Biology and Evolution* 2001; **18**: 1189-1203.
- 460 16 Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C: A genetic landscape
461 reshaped by recent events: Y-chromosomal insights into Central Asia. *American*
462 *Journal of Human Genetics* 2002; **71**: 466-482.
- 463 17 Ausubel FM, Brent R, Kingston RE *et al*: Current Protocols in Molecular Biology.
464 New York, 2001.
- 465 18 Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning. A Laboratory Manual. New
466 York, Cold Spring Harbor, 1982.
- 467 19 Segurel L, Martinez-Cruz B, Quintana-Murci L *et al*: Sex-specific genetic structure
468 and social organization in Central Asia: insights from a multi-locus study. *PLoS Genet*
469 2008; **4**: e1000200.

- 470 20 Rosenberg NA, Pritchard JK, Weber JL *et al*: Genetic structure of human populations.
471 *Science* 2002; **298**: 2381-2385.
- 472 21 Cann HM, de Toma C, Cazes L *et al*: A human genome diversity cell line panel.
473 *Science* 2002; **296**: 261-262.
- 474 22 Zhivotovsky LA, Rosenberg NA, Feldman MW: Features of evolution and expansion
475 of modern humans, inferred from genomewide microsatellite markers. *American*
476 *Journal of Human Genetics* 2003; **72**: 1171-1186.
- 477 23 ElMousadik A, Petit RJ: High level of genetic differentiation for allelic richness
478 among populations of the argan tree *Argania spinosa* (L) Skeels endemic to Morocco.
479 *Theoretical and Applied Genetics* 1996; **92**: 832-839.
- 480 24 Goudet J: FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal*
481 *of Heredity* 1995; **86**: 485-486.
- 482 25 Nei M: Estimation of Average Heterozygosity and Genetic Distance from a Small
483 Number of Individuals. *Genetics* 1978; **89**: 583-590.
- 484 26 Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F: GENETIX 4.05, logiciel
485 sous Windows TM pour la génétique des populations. Laboratoire Génome,
486 Populations, Interactions, CRNS UMS 5171, Université de Montpellier II, Montpellier
487 (France) 1996-2004.
- 488 27 Inc. SI: JMP Statistics and Graphics Guide, Version 5.1. Cary, NC: SAS Institute Inc.
489 2003.
- 490 28 Rousset F: GENEPOP '007: a complete re-implementation of the GENEPOP software
491 for Windows and Linux. *Molecular Ecology Resources* 2008; **8**: 103-106.
- 492 29 Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from
493 metric distances among DNA haplotypes - Application to human mitochondrial -DNA
494 restriction data *Genetics* 1992; **131**: 479-491.

- 495 30 Excoffier L, Laval LG, Schneider S: Arlequin ver. 3.0: An integrated software
496 package for population genetics data analysis. *Evolutionary Bioinformatics Online*
497 2005; **1**: 47-50.
- 498 31 Ray N: PATHMATRIX: a geographical information system tool to compute effective
499 distances among samples. *Molecular Ecology Notes* 2005; **5**: 177-180.
- 500 32 Pritchard JK, Stephens M, Donnelly P: Inference of population structure using
501 multilocus genotype data. *Genetics* 2000; **155**: 945-959.
- 502 33 Rosenberg NA: Standardized subsets of the HGDP-CEPH human genome diversity
503 cell line panel, accounting for atypical and duplicated samples and pairs of close
504 relatives. *Annals of Human Genetics* 2006; **70**: 841-847.
- 505 34 Rosenberg NA, Mahajan S, Gonzalez-Quevedo C *et al*: Low levels of genetic
506 divergence across geographically and linguistically diverse populations from India.
507 *Plos Genetics* 2006; **2**: 2052-2061.
- 508 35 Hubisz MJ, Falush D, Stephens M, Pritchard JK: Inferring weak population structure
509 with the assistance of sample group information. *Molecular Ecology Resources* 2009;
510 **9**: 1322-1332.
- 511 36 Jakobsson M, Rosenberg NA: CLUMPP: a cluster matching and permutation program
512 for dealing with label switching and multimodality in analysis of population structure.
513 *Bioinformatics* 2007; **23**: 1801-1806.
- 514 37 Falush D, Stephens M, Pritchard JK: Inference of population structure using
515 multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*
516 2003; **164**: 1567-1587.
- 517 38 Wang JL: Maximum-likelihood estimation of admixture proportions from genetic
518 data. *Genetics* 2003; **164**: 747-765.

- 519 39 Belle EMS, Landry PA, Barbujani G: Origins and evolution of the Europeans'
520 genome: evidence from multiple microsatellite loci. *Proceedings of the Royal Society*
521 *B-Biological Sciences* 2006; **273**: 1595-1602.
- 522 40 Qamar R, Ayub Q, Mohyuddin A *et al*: Y-chromosomal DNA variation in Pakistan.
523 *American Journal of Human Genetics* 2002; **70**: 1107-1124.
- 524 41 Zerjal T, Xue YL, Bertorelle G *et al*: The genetic legacy of the mongols. *American*
525 *Journal of Human Genetics* 2003; **72**: 717-721.
- 526 42 Bosch E, Calafell F, Gonzalez-Neira A *et al*: Paternal and maternal lineages in the
527 Balkans show a homogeneous landscape over linguistic barriers, except for the
528 isolated Aromuns. *Annals of Human Genetics* 2006; **70**: 459-487.
- 529 43 Manica A, Prugnolle F, Balloux F: Geography is a better determinant of human
530 genetic differentiation than ethnicity. *Human Genetics* 2005; **118**: 366-371.
- 531 44 Brunet F: La Néolithisation en Asie Centrale: un état de la question. *Paléorient* 1999;
532 **24**: 27-48.
- 533 45 Li H, Cho K, Kidd JR, Kidd KK: Genetic Landscape of Eurasia and "Admixture" in
534 Uyghurs. *American Journal of Human Genetics* 2009; **85**: 934-937.
- 535 46 Heyer E, Balaesque P, Jobling MA *et al*: Genetic diversity and the emergence of
536 ethnic groups in Central Asia. *Bmc Genetics* 2009; **10**: 8.
- 537 47 Soucek S. A history of Inner Asia. Cambridge, Cambridge University Press, 2000.
- 538 48 Li JZ, Absher DM, Tang H *et al*: Worldwide human relationships inferred from
539 genome-wide patterns of variation. *Science* 2008; **319**: 1100-1104.
- 540 49 Dupaine B: L'artisanat Hazâra; in CEREDAF: Paysage du centre de l'Afghanistan -
541 Paysages Naturels, paysages culturels. Paris, 2010, pp 212-222.
- 542
543

544 **Figure legends**

545

546 **Figure 1** Geographic location of the 26 Central Asian populations sampled. Linguistic
547 affiliation, as well as admixture proportions from putative parental origins (Central/South
548 Asia, East Asia, Europe and Middle East) are also indicated. See Table 1 for acronyms.

549

550 **Figure 2** Correspondence analysis (CA) based on the table of allele counts in Central Asia
551 (a). The first two factorial components (FC) are represented, and their relative contribution to
552 the total inertia are indicated. Colors indicate language affiliation; blue: Indo-Iranian
553 speakers; orange: Turkic speakers. CA based on the table of allele counts in Eurasian
554 populations (b). Colors represent major geographic regions; purple: Europe; grey: Middle
555 East; green: Central/South Asia; red: East Asia.

556

557 **Figure 3** Population structure inferred from microsatellite data using the software package
558 STRUCTURE. K represents the number of putative clusters. Each individual is represented by
559 a vertical line, divided into up to K colored segments, each of which represents the
560 individual's estimated membership fraction to that cluster. Each output represents the matrix
561 of membership coefficients averaged over 40 independent runs with CLUMPP. The data
562 consisted in 767 individuals from 26 Central Asian populations genotyped at 27 microsatellite
563 loci, plus 869 individuals from 44 African and Eurasian populations from the HGDP-CEPH
564 Human Genome Diversity Cell Line Panel. See Table 1 for acronyms.

Table 1 Description of the 26 Central Asian studied populations

Sampled populations (area)	Acronym	Location	Language family	Long.	Lat.	<i>n</i>
Tajiks (Samarkand)	TJA	Uzbekistan / Tajikistan border	Indo-Iranian	39.54	66.89	31
Tajiks (Samarkand)	TJU	Uzbekistan / Tajikistan border	Indo-Iranian	39.50	67.27	29
Tajiks (Ferghana)	TJR	Tajikistan / Kyrgyzstan border	Indo-Iranian	40.36	71.28	29
Tajiks (Ferghana)	TJK	Tajikistan / Kyrgyzstan border	Indo-Iranian	40.25	71.87	26
Tajiks (Gharm)	TJE	Northern Tajikistan	Indo-Iranian	39.12	70.67	25
Tajiks (Gharm)	TJN	Northern Tajikistan	Indo-Iranian	38.09	68.81	24
Tajiks (Gharm)	TJT	Northern Tajikistan	Indo-Iranian	39.11	70.86	25
Tajiks (Penjikent)	TDS	Uzbekistan / Tajikistan border	Indo-Iranian	39.28	67.81	25
Tajiks (Penjikent)	TDU	Uzbekistan / Tajikistan border	Indo-Iranian	39.44	68.26	25
Tajiks (Yagnobs from Dushanbe)	TJY	Western Tajikistan	Indo-Iranian	38.57	68.78	25
Uzbeks (Ferghana)	UZA	Uzbekistan / Kyrgyzstan border	Turkic	40.77	72.31	25
Uzbeks (Penjikent)	UZT	Northern Tajikistan	Turkic	39.49	67.54	25
Uzbeks (Bukhara)	LUZn	Central Uzbekistan	Indo-Iranian	39.70	64.38	20
Uzbeks (Bukhara)	LUZa	Central Uzbekistan	Indo-Iranian	39.73	64.27	20

Uzbeks (Karakalpakia)	UZB	Western Uzbekistan	Turkic	43.04	58.84	35
Karakalpaks (Qongrat from Karakalpakia)	KKK	Western Uzbekistan	Turkic	43.77	59.02	45
Karakalpaks (On Tört Uruw from Karakalpakia)	OTU	Western Uzbekistan	Turkic	42.94	59.78	45
Kazaks (Karakalpakia)	KAZ	Western Uzbekistan	Turkic	43.04	58.84	49
Kazaks (Bukhara)	LKZ	Central Uzbekistan	Turkic	40.08	63.56	25
Kyrgyz (Andijan)	KRA	Uzbekistan / Kyrgyzstan border	Turkic	40.77	72.31	45
Kyrgyz (Narin)	KRG	Eastern Kyrgyzstan	Turkic	41.60	75.80	18
Kyrgyz (Narin)	KRM	Eastern Kyrgyzstan	Turkic	41.45	76.22	21
Kyrgyz (Narin)	KRL	Eastern Kyrgyzstan	Turkic	41.36	75.50	22
Kyrgyz (Narin)	KRB	Eastern Kyrgyzstan	Turkic	41.25	76.00	24
Kyrgyz (Issyk Kul)	KRT	Eastern Kyrgyzstan	Turkic	42.16	77.57	37
Turkmen (Karakalpakia)	TUR	Western Uzbekistan	Turkic	41.55	60.63	47

Long., longitude; Lat., latitude. *n*, sample size.

1 **Table 2** Genetic diversity in the studied populations and in Eurasia

World Area	Population	AR	H_e
Central Asia	KAZ	7.9	0.784
Central Asia	KKK	7.8	0.782
Central Asia	KRA	7.5	0.769
Central Asia	KRB	7.3	0.757
Central Asia	KRG	7.7	0.779
Central Asia	KRL	7.8	0.778
Central Asia	KRM	7.6	0.752
Central Asia	KRT	7.7	0.761
Central Asia	LKZ	7.8	0.778
Central Asia	LUZa	8.3	0.817
Central Asia	LUZn	8.6	0.821
Central Asia	OTU	8.0	0.784
Central Asia	TDS	7.7	0.784
Central Asia	TDU	7.9	0.805
Central Asia	TJA	8.0	0.806
Central Asia	TJE	8.4	0.814
Central Asia	TJK	8.6	0.820
Central Asia	TJN	8.4	0.811
Central Asia	TJR	8.6	0.812
Central Asia	TJT	8.5	0.812
Central Asia	TJU	8.5	0.811
Central Asia	TJY	7.9	0.799
Central Asia	TUR	8.5	0.812

Central Asia	UZA	9.0	0.817
Central Asia	UZB	8.5	0.774
Central Asia	UZT	8.4	0.795
Central Asia (pooled populations)		12.58	0.803
Central/South Asia		12.66	0.819
East Asia		11.4	0.705
Europe		11.83	0.808
Middle East		13.17	0.827

2 *AR*, allelic richness; H_e , expected heterozygosity. *AR* was calculated using a common sample
3 size of $n = 13$ diploid individuals for the Central Asian samples, and a common samples size
4 of $n = 123$ diploid individuals for the regional samples. These sample sizes correspond to the
5 smallest number of genes sampled at a locus, including missing data.

6 **Table 3** AMOVA of the 26 Central Asian studied populations

Grouping	Source of variation	Percentage of variation	F_{ST}	F_{SC}	F_{CT}
Linguistic affiliation	Among groups	1.09			0.010***
	Among populations within groups	0.91		0.009***	
	Within populations	98.0	0.020***		
Ethnicity	Among groups	0.69			0.007***
	Among populations within groups	0.91		0.009***	
	Within populations	98.39	0.016***		

7 * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$.

8

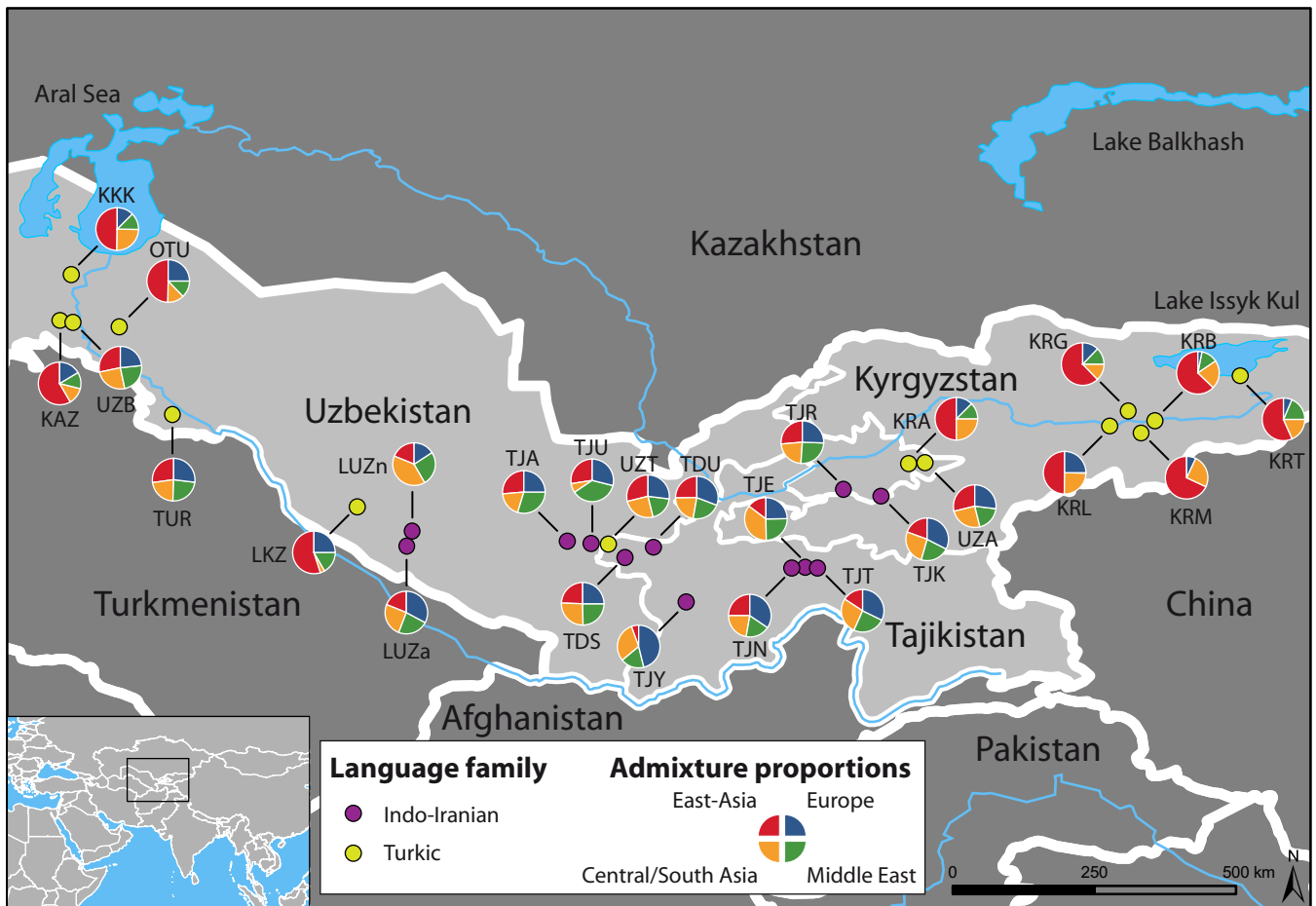
9 **Table 4** Maximum-likelihood estimates of admixture proportions in the 26 Central Asian populations

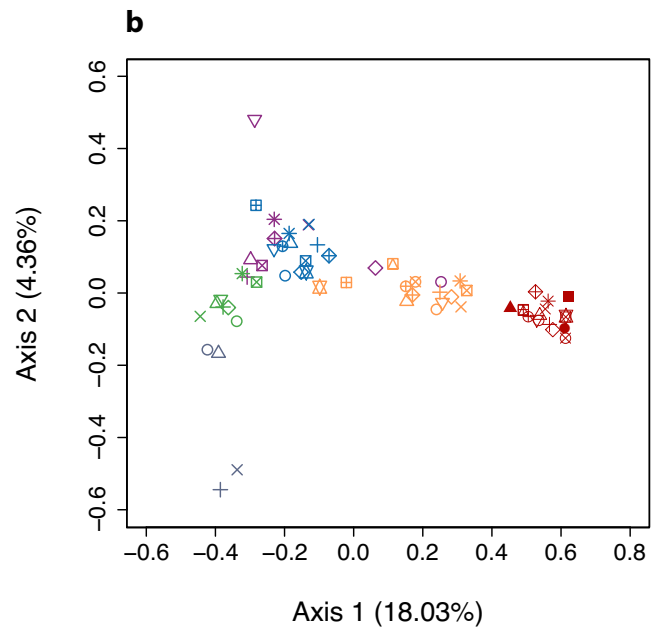
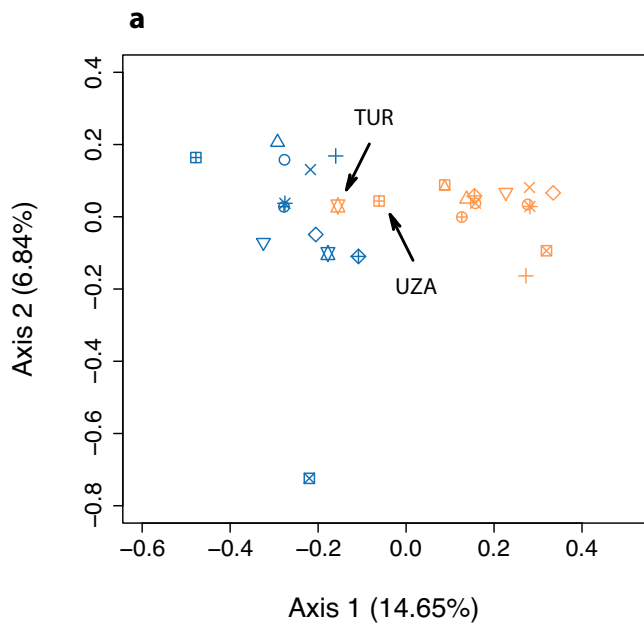
Population	Ethnic group	Putative parental group			
		Europe	Middle East	Central/South Asia	East Asia
KAZ	Kazakh	0.166	0.125	0.126	0.583
LKZ	Kazakh	0.252	0.166	0.033	0.549
KKK	Karakalpak	0.126	0.127	0.250	0.497
OTU	Karakalpak	0.250	0.128	0.125	0.497
KRA	Kyrgyz	0.125	0.126	0.250	0.499
KRB	Kyrgyz	0.031	0.125	0.218	0.625
KRG	Kyrgyz	0.124	0.126	0.129	0.621
KRL	Kyrgyz	0.250	0.004	0.250	0.495
KRM	Kyrgyz	0.072	0.000	0.250	0.678
KRT	Kyrgyz	0.066	0.184	0.184	0.566
TUR	Turkmen	0.271	0.236	0.221	0.272
UZA	Uzbek	0.271	0.192	0.250	0.286

UZB	Uzbek	0.235	0.234	0.250	0.281
UZT	Uzbek	0.271	0.192	0.250	0.287
LUZa	Uzbek	0.330	0.229	0.254	0.187
LUZn	Uzbek	0.160	0.257	0.395	0.188
TDS	Tajik	0.250	0.249	0.258	0.242
TDU	Tajik	0.310	0.219	0.220	0.251
TJA	Tajik	0.250	0.298	0.190	0.262
TJE	Tajik	0.250	0.248	0.358	0.145
TJK	Tajik	0.327	0.219	0.260	0.194
TJN	Tajik	0.345	0.184	0.221	0.250
TJR	Tajik	0.256	0.256	0.226	0.262
TJT	Tajik	0.324	0.244	0.274	0.158
TJU	Tajik	0.290	0.366	0.071	0.273
TJY	Tajik	0.462	0.179	0.303	0.055

10
11 Shaded cells correspond to Turkic-speaking populations, and non-shaded cells to Indo-Iranian-speakers.

12
13





Central Asia
(Indo-Iranian language)

- LUZa ▽ TJE ✕ TJU
- △ LUZn ▣ TJK ▢ TJY
- + TDS * TJN
- × TDU ◆ TJR
- ◇ TJA ● TJT

Central Asia
(Turkic language)

- KAZ ▽ KRL ✕ TUR
- △ KKK ▣ KRM ▢ UZA
- + KRA * KRT ✖ UZB
- × KRB ◆ LKZ ▣ UZT
- ◇ KRG ● OTU

Central/South Asia

- Uygur ▽ Kalash
- △ Balochi ▣ Makrani
- + Brahui * Pathan
- × Burusho ◆ Sindhi
- ◇ Hazara

Europe

- Basque ▽ Orcadian
- △ French ▣ Russian
- + Bergamo * Adygei
- × Sardinian
- ◇ Tuscan

Middle East

- Druze
- △ Palestinian
- + Bedouin
- × Mozabite

East Asia

- Cambodian ▽ Lahu ✕ She ● Japanese
- △ Dai ▣ Miaozi ▢ Tu ■ Yakut
- + Daur * Mongola ✖ Tujia
- × Han ◆ Naxi ▣ Xibo
- ◇ Hezhen ● Oroqen ▲ Yizu

