Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/3/2/e1601877/DC1

Supplementary Materials for

Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago

Veronika Siska, Eppie Ruth Jones, Sungwon Jeon, Youngjune Bhak, Hak-Min Kim, Yun Sung Cho, Hyunho Kim, Kyusang Lee, Elizaveta Veselovskaya, Tatiana Balueva, Marcos Gallego-Llorente, Michael Hofreiter, Daniel G. Bradley, Anders Eriksson, Ron Pinhasi, Jong Bhak, Andrea Manica

Published 1 February 2017, *Sci. Adv.* **3**, e1601877 (2017)
DOI: 10.1126/sciadv.1601877

This PDF file includes:

- fig. S13. *D* scores of the form *D*(*X*, Khomani; MA1, DevilsGate1), with all modern populations in our panel and selected ancient samples, using all SNPs.
- fig. S14. Outgroup f_3 scores of the form f_3(*X*, Ust'-Ishim; Khomani), with modern populations and selected ancient samples (MA1, Kotias, Loschbour, and Stuttgart), using all SNPs, with $f_3 > 0.15$ displayed.
- fig. S15. *D* scores of the form *D*(*X*, Khomani; Ust'-Ishim, DevilsGate1), with all modern populations in our panel and selected ancient samples, using all SNPs.
- fig. S16. *D* scores of the form *D*(*X*, Khomani; Ust'-Ishim, DevilsGate2), with all modern populations in our panel and selected ancient samples, using all SNPs.
- fig. S17. Comparison of Devil's Gate–related ancestry in the Ulchi and European hunter-gatherer–related ancestry in European populations.
- fig. S18. Comparison of Devil's Gate–related ancestry in the Ulchi and Early European farmer–related ancestry in European populations.
- fig. S19. Comparison of Devil's Gate–related ancestry in the Ulchi and Bronze Age Steppe–related ancestry in European populations.
- table S1. Details of sample preparation and sequencing.
- table S2. mtDNA contamination estimates.
- table S3. Admixture f_3(Source1, Source2; Target) for the Ulchi with $z < -1$ using all SNPs.
- table S4. Admixture f_3(Source1, Source2; Target) for the Ulchi with $z < -1$ using only transversion SNPs.
- table S5. Admixture f_3(Source1, Source2; Target) for the Sardinians using all SNPs and showing the 10 most significantly negative pairs.
- table S6. Admixture f_3(Source1, Source2; Target) for the Lithuanians using all SNPs and showing the 10 most significantly negative pairs.
- Legends for extended data figs. S1 to S8
- Legends for extended data tables S1 to S9
- References (*55–84*)

Other Supplementary Material for this manuscript includes the following:
(available at advances.sciencemag.org/cgi/content/full/3/2/e1601877/DC1)

- extended data fig. S1 (.pdf format). Results from ADMIXTURE analysis using the regional panel, all SNPs, and MapDamage treatment on samples from Devil's Gate and setting the number of clusters to $K = 2$ to 10.
- extended data fig. S2 (.pdf format). Results from ADMIXTURE analysis using the regional panel, transversion SNPs, and MapDamage treatment on samples from Devil's Gate and setting the number of clusters to $K = 2$ to 10.
- extended data fig. S3 (.pdf format). Results from ADMIXTURE analysis using the regional panel, all SNPs, and no MapDamage treatment on samples from Devil's Gate and setting the number of clusters to $K = 2$ to 10.
- extended data fig. S4 (.pdf format). Results from ADMIXTURE analysis using the regional panel, transversion SNPs, and no MapDamage treatment on samples from Devil's Gate and setting the number of clusters to $K = 2$ to 10.

- extended data fig. S5 (.pdf format). Results from ADMIXTURE analysis using the total panel, all SNPs, and MapDamage treatment on samples from Devil's Gate and setting the number of clusters to K = 2 to 20.
- extended data fig. S6 (.pdf format). Results from ADMIXTURE analysis using the total panel, transversion SNPs, and MapDamage treatment on samples from Devil's Gate and setting the number of clusters to K = 2 to 20.
- extended data fig. S7 (.pdf format). Results from ADMIXTURE analysis using the total panel, all SNPs, and no MapDamage treatment on samples from Devil's Gate and setting the number of clusters to K = 2 to 20.
- extended data fig. S8 (.pdf format). Results from ADMIXTURE analysis using the total panel, transversion SNPs, and no MapDamage treatment on samples from Devil's Gate and setting the number of clusters to K = 2 to 20.
- extended data table S1 (Microsoft Excel format). Sample information.
- extended data table S2 (Microsoft Excel format). ADMIXTURE proportions.
- extended data table S3 (Microsoft Excel format). Outgroup f_3 statistics for Devil's Gate.
- extended data table S4 (Microsoft Excel format). Outgroup f_3 and space.
- extended data table S5 (Microsoft Excel format). Outgroup f_3 for MA1 and Ust'-Ishim.
- extended data table S6 (Microsoft Excel format). D scores for MA1 and Ust'-Ishim.
- extended data table S7 (Microsoft Excel format). D scores for the Ulchi.
- extended data table S8. Admixture f_3 for the Koreans and the Japanese.
- extended data table S9 (Microsoft Excel format). Phenotypes of interest.

SUPPLEMENTARY MATERIALS AND METHODS

1. Archaeology

Devil's Gate (Chertovy Vorota) is a karst cave situated in the mountainous part of Primorye Province, 12 km from the town Dalnegorsk (Primorsky Krai, Russian Far East), 660 meters above sea level and ~35m above the Krivaya River. The cave was first excavated in 1973 by a Soviet team under the directorship of Tatarnikov V.A. (*55*). This is the only known Neolithic cave site in Primorye (*56*) .

The large cave has a rectangular entrance and consists of a 45m long and 10m wide gallery of which the total area excavated was ~200 m^2. Most of the cave's area was occupied as a dwelling space (*56*). The site contains a single occupational layer which belongs to the 'Rudninskaya' Neolithic culture. In the central part of the hall, excavations revealed the remains of a rectangular pit dwelling around 45 m^2 in area which was walled by wooden poles (*56*).

Cultural material recovered included stone, bone and antler tools, pottery, bone and shell ornaments, as well as good preservation of other organics including wood and textile artefacts (*10*). The lithic assemblage includes retouched scrapers, arrows, knives, drills, and other tools. The bone tools include needles of various sizes, harpoons, and bases of insert tools. The composition of the tool assemblages is typical and indicative of a subsistence which is based on hunting and fishing (*56*).

Excavations also yielded 14 clay vessels made of local materials and a large number of shards. The pottery is flat-bottomed and conical-shaped with comb pattern decoration (*10*); the method of construction was coiling. The vessels were decorated by stamping and appliqué.

Prior to radiometric dating, the cultural layer at the cave was indirectly assigned to the 5^{th} millennium BC based on the artefacts, which are similar to those recovered from a settlement situated near the Rudnaya River in the Dalnegorsk district (*57*).

The excavations yielded fragments of textiles, fishing nets, cords and mats made of plant fabrics and wooden artefacts made of birch bark. A large number of bones of animals, acorns, fruits and birch bark were found in the center (*57*). The assemblage also includes

numerous ornaments including bone pendants from boar canines and beads made of shell, bone and stone.

Animal bones belong mostly to brown and black bears, wild boar, red deer, and badger. The fauna also comprises of the remains of birds (grouse, ptarmigan, dove and duck) and freshwater and anadromous fish (salmon) (*55*). The assemblage does not include any domestic plant or faunal remains which suggests that the inhabitants of the site belonged to a hunter-gatherer-fisher population who manufactured utilitarian low-quality domestic pottery.

2. Osteology and radiocarbon dates

The fragmented remains of seven individuals were unearthed from the floor of the dwelling, none of which were recovered in their primary burial anatomical position.

- Skull A (DevilsGate4) – juvenile, 6-7 years (genetically determined to be female)
- Skull Д (DevilsGate3) – female, 50-60 years old
- Skull Б (DevilsGate2) – male, 20-25 years old
- Skull E (DevilsGate1) – female, 40-50 years old
- Skull Ж (DevilsGate5) – complete skull of a young person, 18-20 years old (believed to be a male based on morphology but genetically determined to be female)
- Skull В – sub-adult, 12-13 years old (no aDNA analysis)
- Skull Г – male, ~50 years old (no aDNA analysis)

Complete and fragmented skulls were recovered from four of these individuals, three adults and one subadult (12-13 years of age) (*58*). Of the three adult skulls, two were believed to be from males (although one of these was genetically determined to be a female, see Supp. S.7.) and one of a female.

Direct AMS radiocarbon dates were obtained for two of the individuals. Specimen Б (DevilsGate2), was dated at the Oxford AMS Radiocarbon Laboratory to 6765±40 uncalibrated before present (OxA 27677) and specimen E (DevilsGate1) was dated to

6756±37 uncal. bp (OxA 27678). The calibrated range of both dates is 5726-5622 cal BC (2 SD. 95.4% confidence interval range, fig. S1).

The skulls all have a short, high and wide cranial vault, a wide, high and flat face, with a slight protrusion of the nasal bones and with narrow orbits. The mandible is very wide and is characteristic of hunter-gatherers (*58*). The craniometric features are in general agreement with the genetics in terms of suggesting some level of regional continuity between the Devil's Gate individuals and later populations in the Russian Far East.

3. Authenticity of results

Patterns of molecular damage and the length distribution of reads were assessed using all reads for DevilsGate3, DevilsGate4, and DevilsGate5. As a portion of the reads from DevilsGate1 and DevilsGate2 derived from 50 bp single end sequencing, only reads sequenced with 150 bp paired end sequencing were considered in the following analyses in order to avoid using truncated reads (library ID MOS5A.E1 for DevilsGate1 and MOS4A.E1 for DevilsGate2). MapDamage 2.0 (*44*) was used to assess patterns of molecular damage which are typical of ancient DNA. Only reads with mapping quality $\geqslant$20 were considered. An increased rate (up to 11%) of C to T misincorporations was found at the 5' ends of reads with reciprocal patterns of G to A misincorporations at the 3' read termini (fig. S2). The sequence length distribution was analysed as in (*59*). For all samples a peak in DNA sequence length is visible at <70 bp, consistent with the short fragment length expected for ancient molecules (*60*) (fig. S3).

Low coverage data like ours often do not provide sufficient information to distinguish the C to T mutations on the sample relative to the reference sequence, which appear close to the 5' terminus, from damage. Such positions can be downscaled in quality and dropped, which can, however, lead to a bias pulling the ancient sample closer to the human reference genome if substitutions that are not caused by damage are dropped by this procedure. On the other hand, not applying MapDamage can lead to the opposite bias due to damage. Thus, we also replicated our analysis on versions of Devil's Gate samples without MapDamage rescaling, to confirm that neither of these biases affect our conclusions.

4. Contamination estimates

The rate of mitochondrial contamination was assessed for our highest coverage samples, DevilsGate1 and DevilsGate2. This was calculated by evaluating the percentage of non-consensus bases at haplogroup-defining positions (haplogroup D4 for DevilsGate1 and M for DevilsGate2) using bases with quality $\geqslant$ 20 (*17, 20*). Low contamination rates of 0.87% (95% CI 0.28%-2.37%) and 0.59% (95% CI 0.03-3.753%) were found for DevilsGate1 and DevilsGate2, respectively, when all bases were considered (see table S2 for more details). This was reduced to 0% for both samples when non-consensus base calls, which could be explained by deamination, were omitted from the analysis.

We also used schmutzi, a tool which employs a Bayesian maximum a posteriori algorithm (*12*), to estimate the mitochondrial contamination for DevilsGate1 (we did not estimate the contamination rate for DevilsGate2 as this tool estimates a contamination prior using deamination frequencies at read termini and as 99% of reads from DevilsGate2 derive from 50 bp sequencing, its read termini are likely truncated). For DevilsGate1 only reads from $\geqslant$ 100bp sequencing were used in the analysis. The contDeam.pl script was run using the --library double option followed by the schmutzi.pl script with default parameters, using a dataset of putative contaminants provided with the schmutzi package. We found contamination estimates of 1% (95% CI 0-2%), in line with our previous estimates. It should be noted that this program will underestimate the contamination rate if the contaminating molecules are deaminated or if there are multiple contaminating sources.

Since other ways of estimating contamination were not possible (e.g. based on the X-chromosome), we also attempted to replicate our main results using only reads with evidence of postmortem damage. We applied PMDtools (9), a framework assigning likelihood scores to degraded sequences that are unlikely to originate from modern contamination. In this part of the analysis, we restricted our reads to those with a PMD score of at least 3. This greatly decreased our coverage, to 0.0050X for DevilsGate1 and 0.0012X for DevilsGate2, but it was still enough to perform two of our most robust analysis: the outgroup f_3 statistics (fig. S4) and a Principal Component Analysis (figs. S5 and S6 using the world-wide and the regional panel, respectively). The results were considerably noisier, but looked qualitatively similar to the results obtained using all reads: the samples look Northeast Asian, closest to the Ulchi.

5. Mitochondrial Haplogroup Determination

Mitochondrial consensus sequences were generated for DevilsGate1 and DevilsGate2 using ANGSD (*46*). Called positions were required to have a depth of coverage ⩾ 3 and only bases with quality ⩾ 20 were considered. The resulting FASTA files were uploaded to HAPLOFIND (*47*) for haplogroup determination. Mutations defining the assigned haplogroup were also manually checked. Coverage was calculated using GATK DepthOfCoverage (*43*).

DevilsGate1 (9.76-fold coverage) was assigned to haplogroup D4. DevilsGate1 has one of the two mutations which define haplogroup D (T16362C, 13-fold coverage; the other position had no spanning sequence data) and all three mutations which are associated with haplogroup D4 (G3010A, 2-fold coverage; C8414T, 3-fold coverage; C14668T, 6-fold coverage).

DevilsGate2 (2.75-fold coverage) was assigned to haplogroup M. DevilsGate2 had three of the four mutations which define haplogroup M (T489C, 7-fold coverage; C10400T, 3-fold coverage; T14783C, 2 -fold coverage). DevilsGate2 was assigned with equal likelihood to subhaplogroups M9 and M13'46'61, suggesting that this sample may not be accurately assigned to a subhaplogroup due to low sequence coverage.

6. Sex determination

Sex was assigned using the script described in (*48*). The observed fraction of Y chromosome alignments compared to the total number of alignments to the X and Y chromosome and its estimated 95% confidence interval was R_Y = 0.0057 (0.0053-0.0061) and R_Y = 0.0059 (0.0054-0.0064) for DevilsGate1 and DevilsGate2, respectively, implying that both of them were females. DevilsGate1's sex is in accordance with that inferred from archaeology, but DevilsGate2 was thought to be male based on its skull.

7. Affinities of the samples from Devil's Gate

1. Principal Component Analysis

In order to explore where our samples from Devil's Gate are placed in the context of the main axes defining modern genetic variation, we performed a Principal Component Analysis (PCA) with two different reference panels, both subsets of the worldwide panel of contemporary and ancient individuals from (*5*). The analysis was carried out using

EIGENSOFT 6.0.1 smartpca (*18*) with the lsqproject and normalisation options on, the outlier removal option off and one SNP from each pair in linkage disequilibrium with r^2>0.2 removed. Ancient samples were projected onto the Principal Components defined by modern populations.

We investigated two different reference panels, the first consisting of all modern individuals in our worldwide set of populations. Here, the Devil's Gate samples clearly clustered with northern Asian populations, in particular, with those from East Asia, Central Asia and Siberia, referred to as the "regional panel" (fig. S5). This suggests a degree of continuity in North-Eastern Asia, with our 7.7 ky old sample falling broadly within modern variation.

Next, we focused our attention on the regional panel of East Asian, Central Asian and Siberian populations that were close to our sample on the broader panel. Both Devil's Gate individuals clustered with populations from the Amur Basin, particularly close to the Ulchi, Oroqen, and Hezhen (fig. S6).

Results based on transversion SNPs are shown on fig. S7 for the global and fig. S8 for the regional panel.

2. ADMIXTURE analysis

A clustering analysis was performed using ADMIXTURE version 1.23 (*19*). SNPs in linkage disequilibrium were thinned using PLINK 1.07 with parameters –indep-pairwise 200 25 0.5 resulting in a set of 334,359 SNPs for analysis (91,379 transversions). K=2-20 clusters were explored for the global panel and K=1-10 for the regional panel, using 10 independent runs with fivefold cross-validation at each K with different random seeds. The minimal cross-validation error was found at K=18 for the global panel (fig. S9) and K=5 for the regional (East & Central Asian) panel (fig. S10), but the error already started plateauing around K=9 for the global panel, suggesting little improvement beyond this value. Furthermore, results for the regional panel were largely similar to those from the global panel for East Asian populations.

For both panels, the main clusters reported in previous analyses were found (*3–5*). For every K value for the global and K≤7 for the regional panel), Devil's Gate consisted of two components in roughly equal proportions. The first is an East Asian component, prevalent in the Han Chinese, Koreans and Japanese, among others, and the second is a Siberian component, typical for the Nganasan. Other populations from the Amur Basin, Koreans and

Japanese also had the same components, but with a higher proportion of "East Asian" ancestry and a small amount of Taiwanese-like component for the Koreans and Japanese.

From K=8 (regional panel) onwards, a new component corresponding to Devil's Gate appeared. It appeared in various other East Asian populations, including those from the Amur Basin (especially the Ulchi), modern-day China, Japan and Korea.

ADMIXTURE results for both panels, using all SNPs or transversions only, are shown on Extended Data figs. S1-S8.

3. Outgroup f_3 statistics

We used outgroup f_3-statistics to estimate the amount of shared drift between inhabitants of Devil's Gate and a range of modern-day and ancient populations. We computed $f_3(X, \text{DevilsGate}; \text{Khomani})$, where X was a population from our panel, and the Khomani (Khoisan) acted as outgroup. f_3 statistics were calculated with the 3PopTest tool from the AdmixTools (*49*).

When considering our global panel of modern populations, the Ulchi was the closest (about one standard deviation higher than the second closest population), followed by a heterogeneous set of populations from Northeast Eurasia with values close to each other (see Fig. 2 using all SNPs and MapDamage treatment, fig. S4 using PMD-filtered data and extended data table S3 for all other results). These populations include other Tungusic-speaking populations originating from the Amur Basin (Oroqen, Hezhen, Daur, and Xibo), a number of eastern Siberians (Nganasan, Koryak, and Eskimo), the Koreans and the Japanese. They were then followed by variousmore southern East Asian populations, primarily from present-day China and a number of eastern Central Asian populations. South East Asians and western Central Asians scored lower, in agreement with their larger geographic distance from Devil's Gate. The pattern of a high negative correlation between f_3 and geographic distance on land (computed using the graph approach in (*61*) is in agreement with neutral drift leading to a simple isolation-by-distance pattern (extended data table S4).

None of the published ancient samples gave a large outgroup f_3 score with Devil's Gate, an unsurprising result given their distant geographic origin. Generally, they give similar results to the modern population that resembles them genetically (e.g. ancient European samples - modern European populations). This is in agreement with the inhabitants of Devil's Gate belonging to a separate lineage of ancient East Asians.

8. Relation to MA1 and Ust'Ishim

Out of the ancient samples in our panel, MA1 is the only one with ties to eastern Eurasians, northern Siberians in particular (*15*). MA1 had a small contribution to modern populations and is considerably older than our genome. However, the unusually high affinity of Devil's Gate to Native Americans (the modern populations closest to MA1), given their geographical distance, raises the possibility of a tie between the ancient North-Eurasian lineage that MA1 belongs to and that of Devil's Gate.

1. MA1

1. Outgroup f_3 with MA1

First, we used the outgroup f_3-statistics to investigate if the inhabitants of Devil's Gate were related to the ancient Siberian lineage represented by MA1 that contributed to modern-day Native Americans and, to a lesser extent, to populations from northern Eurasia. We considered all modern-day and ancient populations in our panel, including the two new samples from Devil's Gate. We computed f_3(*X*, MA1; Khomani), where *X* was a population from our panel, and the Khomani (Khoisan) acted as outgroup. f_3 statistics were calculated with the 3PopTest tool from the AdmixTools package (*49*).

Devil's Gate samples did not show any particular closeness to MA1, with outgroup- f_3 scores similar to a wide range of Eurasian populations from the southern parts of Asia without close ties to MA1 (fig. S11; extended data table S5). They are as close to MA1 as modern East Asian populations and further away than many modern-day populations (e.g. Native Americans or even northern Europeans). This is in agreement with MA1 belonging to a distinct lineage, without major contribution to the inhabitants of the Devil's Gate cave or to modern-day East Asians.

2. D statistics with MA1

Next, we used D statistics to investigate how MA1 and Devil's Gate are related to modern populations. We considered all modern-day and ancient populations in our panel, including the two new samples from Devil's Gate. We computed D(X, Khomani; MA1, DevilsGate1), where X was a population from our panel, and the Khomani (Khoisan) acted as outgroup. D-statistics were calculated with the qpDstat tool from the AdmixTools package (*49*).

Our result showed that modern East Asians were significantly closer to Devil's Gate than to MA1. Native Americans and Africans were not significantly closer to either and Europeans

were significantly closer to MA1 than to Devil's Gate (figs. S12 and S13; extended data table s6). MA1 and Devil's Gate did not form a clade against any well-defined group of populations. This is in agreement with Native Americans having both East Asian-like and MA1-like ancestry and Europeans mainly originating from their own linage, with a deep split from East Asians, and only a small amount of MA1-like ancestry, mostly in the northern areas.

2. Ust'Ishim

1. Outgroup f_3 with Ust'Ishim

First, we used the outgroup f_3-statistics to investigate if the inhabitants of Devil's Gate were related to the ancient lineage represented by Ust'Ishim that did not seem to have a high contribution to modern populations, but is slightly more related to modern Asians than to modern Europeans. We considered all modern-day and ancient populations in our panel, including the two new samples from Devil's Gate. We computed f_3(*X*, Ust'Ishim; Khomani), where *X* was a population from our panel, and the Khomani (Khoisan) acted as outgroup. f_3 statistics were calculated with the 3PopTest tool from the AdmixTools package (*49*).

Devil's Gate samples did not show any particular closeness to Ust'Ishim, with outgroup- f_3 scores similar to a wide range of Eurasian populations without close ties to MA1 from the southern parts of Asia (all Eurasian populations are within 2 standard deviations of DevilsGate1's and DevilsGate2's scores on fig. S14; extended data table S5). This is in agreement with Ust'Ishim belonging to a distinct lineage, without major contribution to the inhabitants of the Devil's Gate cave or to modern-day East Asians.

2. D statistics with Ust'Ishim

Next, we used D statistics to investigate how Ust'Ishim and Devil's Gate are related to modern populations. We considered all modern-day and ancient populations in our panel, including the two new samples from Devil's Gate. We computed D(X, Khomani; Ust'Ishim, DevilsGate1), where X was a population from our panel, and the Khomani (Khoisan) acted as outgroup. D-statistics were calculated with the qpDstat tool from the AdmixTools package (*49*).

Our result showed that modern East Asians were significantly closer to Devil's Gate than to Ust'Ishim, while Native Americans and Africans and Europeans were not significantly closer to either (figs. S15 and S16; extended data table S6). This is again in agreement both

with Ust'Ishim belonging to a distinct lineage, without major contribution to the inhabitants of the Devil's Gate cave or to modern-day East Asians and Devils's Gate belonging to the Asian branch of modern humans that split from African and European populations around the time or before Ust'Ishim lived.

9. Searching for signals of admixture in the Ulchi

The Ulchi population stood out as the closest population to Devil's Gate based on the outgroup f_3 statistics; however, this alone does not exclude a possible later admixture.

1. Admixture f_3 shows no signal of admixture

In order to search for signals of admixture in the Ulchi, we first used the admixture f_3-statistics in the form $f_3(X,Y;\text{Ulchi})$, using the pq3Pop tools from the AdmixTools package (*49*). We scanned every possible pair of populations *X* and *Y*, taken from our global panel of modern and ancient populations and Devil's Gate. We only considered pairs with at least 1000 SNPs in common. This statistic can be significantly negative if the target population (in this case, the Ulchi) has genetic material from both populations X and Y. When considering all SNPs, all modern pairs gave a positive score, with only a few ancient genomes giving a non-significant (Z>-2) negative score. When considering only transversions, the scores became noisier and some slightly negative scores appeared even for modern populations paired with Devil's Gate samples, but none of these were significant (Z>-2). Pairs of populations with Z<-1 and at least 1000 SNPs covered are shown in tables S3 and S4, using all SNPs or transversions only. These results imply no recent admixture of an appreciable scale. In contrast, European populations gave significantly negative f_3 statistics with numerous pairs of populations, even using our smallest SNP panel that consisted of locations called in DevilsGate2. The 10 pairs of populations with the lowest admixture f_3 values are shown for the Lithuanians (table S6), who harbour the highest hunter-gatherer component on ADMIXTURE and the Sardinians (table S5), who are closest to Early European Farmers on numerous analyses.

2. D statistics cannot reject a breach of continuity

We then tested if the samples from Devil's Gate and the Ulchi form a clade against other populations by examining D statistics of the form D(Khomani, X; Ulchi, Devil's Gate). This statistic deviates from zero if the Ulchi or Devil's Gate are not symmetrically related to population X when compared to an outgroup population (we used the African Khomani San).

For both DevilsGate1 and DevilsGate2, several populations gave results significantly different from zero (|Z|>2; extended data table S7). Populations from Central Asia scored highest, but for DevilsGate1 on the total SNP panel (where we have the most power), a very wide range of populations from all over the world also gave less, but still significantly non-zero scores, in the direction indicating that they are closer to the Ulchi than to Devil's Gate. The high-scoring populations are in agreement with the Central Asian-related ADMIXTURE ancestry components in the Ulchi, but the wide variety of the rest of the populations is difficult to explain. The result could be due to the large uncertainty in SNP calling due to our very low coverage (even the SNPs covered are also covered by a single read in most cases), as well as errors from DNA degradation.

3. Comparison to European populations through proportions from ADMIXTURE

We further used the ADMIXTURE results to compare the levels of Devil's Gate-related ancestry in the Ulchi to hunter-gatherer, Early European Farmer and Bronze Age steppe-related ancestries in modern Europeans. To assess significance, we performed bootstrapping by resampling SNPs and individuals (for each population) simultaneously with replacement 100 times. We then inferred the ancestry proportions by numerically maximising the following likelihood function on the bootstrapped SNPs for each bootstrapped individual, following the logic of (*16*, *51*)

$$L(Q,F) = \sum_i \sum_j \left\{ g_{ij} \ln \left[\sum_k q_{ik} f_{kj} \right] + (2 - g_{ij}) \ln \left[\sum_k q_{ik} (1 - f_{kj}) \right] \right\}$$

where g_{ij} is the count of allele 1 (0, 1 or 2) of individual i at SNP j, q_{ik} is the fraction the inferred ancestral population k contributes to individual i's genome and f_{kj} is the frequency of allele 1 at SNP j in the inferred ancestral population k. We used inferred allele frequencies of the ancestral populations, f_{kj}, from K=8 on the regional panel for the Ulchi (where the Devil's Gate-specific cluster first appears) and K=18 on the global panel for European populations (where the cross-validation error is minimized). We used ADMIXTURE runs on all SNPs and MapDamage-treated samples form Devil's Gate, choosing the run with the lowest cross-validation error out of the 10 independent runs per K value. The likelihood was maximised using the minimize function from the scipy python package, with the "L-BFGS-B" method and proportions bounded between [1e-5,1-1e-5] and constrained to sum to unity.

Ancestry proportions inferred by ADMIXTURE for each sample, overlaid with the distributions of population-level mean proportions from bootstrapping, are shown in extended data table S2 and figs. S17 to S19. For reference, we also show results for Early European Farmer- and Bronze Age Steppe-related ancestries for European populations. The difference between population-level mean proportions is significant for most comparisons: population-level mean ancestry proportions of a component in a European population were lower than the mean proportion of Devil's Gate-related ancestry in the Ulchi in all 100 bootstrap replicates ($p<0.01$). The Early European Farmer-related component in the Sardinians is the only exception: 35 out of the 100 bootstrap replicates resulted in a higher ratio of the former than the latter ($p=0.35$). However, a high level of continuity between Early European Farmers and the Sardinians was already observed in other analysis (*3*).

The proportion of Devil's Gate-related ancestry depends on the choice of k for the ADMIXTURE analysis. However, K=8, which we chose (lowest K with a defined Devil's Gate cluster on all versions) results in the smallest Devil's Gate-related ancestry component in the Ulchi out of all runs with a defined Devil's Gate cluster. Therefore, considering a higher number of clusters would only increase the significance of our findings.

10. Dual origin of the Koreans and the Japanese

In order to search for signals of multiple components in the Koreans and the Japanese, we again used the admixture f_3-statistics in the form $f_3(X,Y;\text{Target})$ for the Japanese and the Koreans as target populations, using the pq3Pop tools from the AdmixTools package (*49*). We scanned every possible pair of populations *X* and *Y*, taken from our global panel of modern and ancient populations and Devil's Gate. We only considered pairs with at least 1000 SNPs in common. This statistic can be significantly negative if the target population has genetic material from both populations X and Y (this does not exclude additional populations also contributing to the target population).

In order to investigate the origin of the northern component and whether it is related to the occupants of Devil's Gate, we calculated admixture f_3 statistics of the form $f_3(X,Y;\text{Korean})$ and $f_3(X,Y;\text{Japanese})$. The strongest signal for both Koreans and Japanese came from Devil's Gate paired with various populations from modern-day China and South-East Asia, regardless of whether the total panel was used or only SNPs called in DevilsGate1 or DevilsGate2. These are followed by pairs where the Ulchi take the place of

Devil's Gate as the proxy for the northern origin. Various other pairs of populations also gave significantly negative results, with one population always originating from Eastern Siberia and the other from present-day China or Southeast Asia.

All pairs giving significantly negative admixture f_3 statistics at a significance level of Z<-2 and containing at least 1000 SNPs are included in extended data table S8.

11. Phenotypes of interest

We investigated phenotypes of interest in our highest coverage sample, DevilsGate1, including some loci known to have been under selection in Eurasian populations. Due to the low quality of our samples, we used BEAGLE (*54*) to impute genotypes using a reference panel containing phased genomes from the 1,000 Genomes Project (26 different populations). Following (*17*), GATK Unified genotyper (*43*) was used to call genotype likelihoods at SNP sites in Phase 3 of the 1,000 Genomes Project. Equal likelihoods were set for positions with no spanning sequence data as well as positions where the observed genotype could be explained by deamination (*17*). We imputed at least 1Mb upstream and downstream from the loci of interest using 10 iterations to estimate genotypes at markers with no sequence coverage. The only position covered by a read was rs74653330 SNP on the OCA2 gene. It had 1x coverage of a C allele (the allele we predicted using imputation). A summary of our results is shown in extended data table S9.

To explore pigmentation, we first looked at the genes *SLC45A2* and *SLC24A5*, contributing to light skin colour in modern-day Europeans (*32, 62–66*). Selected genotypes in these two genes (rs16891982 and rs1426654 for *SLC45A2* and *SLC24A5*, respectively) are known to have been under strong positive selection in Europeans and are almost fixed in modern populations in the region. Our results imply that DevilsGate1 was unlikely to have been homozygous for either of the selected alleles, as is typical for non-Europeans. We also investigated the rs12913832 variant of the *HERC2* gene, where a mutation is found in nearly all people with blue eyes. Our sample was likely homozygous for the ancestral allele, implying brown eyes.

We examined a variant of the *LCT* gene (rs4988235), where the derived genotype is associated with lactose tolerance in adult Europeans (*35*). We found that DevilsGate1 was unlikely to possess the derived genotype similar to modern-day East Asian populations.

We also looked at a number of mutations associated with phenotypes in East Asians. We first investigated two mutations within the *OCA2* gene (rs1800414 and rs74653330) that give signals of positive selection (*67–69*) and are wide-spread and associated with decreased melanin levels in East Asians (*70–72*). The results for the variant rs1800414 were inconclusive, but DevilsGate1 likely did not carry the derived variant at rs74653330. We also looked at two loci with a high frequency differentiation between East Asians and non-East Asians and showing signals of positive selection (rs12821256, rs1407995, rs885479 on the KITLG, DCT and SLC24A2 genes, respectively) (*32*) or otherwise associated with pigmentation (rs1042602, rs1834640, rs26722 on the TYR, RPL7AP62/SLC24A5 and SLC24A2 genes, respectively) (*73*). DevilsGate1 was most likely homozygous for the East Asian variant at both rs12821256 and rs1407995 and did not possess the alleles rare in East Asia at rs885479, rs1042602, rs1834640 and rs26722.

We then looked at the *EDAR* gene. This region also shows signals of positive selection in East Asian populations, with the selected allele linked to certain phenotypic changes (e.g. in tooth morphology (*34*), hair thickness (*33, 74*) and sweat gland density (*75*)), also confirmed using a mouse model. The haplotype with the selected allele is very common in East Asian, Northeast Asian and Native American populations and is estimated to have emerged in central China around 30,000 years ago (*75*). Our sample likely carried at least one copy of the derived mutation, giving increased odds of straight, thick hair as well as shovel-shaped incisors.

We also studied the mutation rs671 on *ALDH2*. The derived variant, of intermediate frequencies (roughly up to 50%) in various East Asian populations (*76, 77*), is associated with alcohol flush and increased risk of hangovers, alcoholism and Esophageal Cancer (*36, 78*). Our sample probably did not carry the derived allele.

We then investigated three East Asian phenotypic traits. DevilsGate1 most likely had at least one copy of the risk allele for common salt-sensitive hypertension (*79*) at rs4961 on the ADD1 gene, common in Asia. She was most likely not homozygous for the risk allele at rs1799971 on the OPRM1 gene, associated with increased craving and stronger effects for alcohol (*80*) and for opioids (*81*) in Asian populations, although the results from association studies on this marker are mixed. The derived variant at rs17822931 on the ABCC11 gene is associated with weaker than usual body odour (*82*) and dry ear wax type (*83*) (for the latter, when homozygous) in East Asians and is indicative of east Asian

ancestry. DevilsGate1's genotype could not be determined but she most likely possessed at least one copy of the derived allele.

Finally, we studied eight loci associated with an increased risk for Type 2 diabetes in East Asian populations (rs1535500, rs6017317, rs6467136, rs9470794, rs3786897, rs6815464, rs7041847, rs831571) (*84*). She was likely heterozygous at two loci associated with increased susceptibility to Type 2 diabetes in East Asians (rs1535500 and rs9470794) and had at least one copy of the disease-associated variant at a third location (rs6815464). Our results at rs3786897 were completely uninformative and at other locations, she most likely had at least one copy of the normal variant.

SUPPLEMENTARY FIGURES

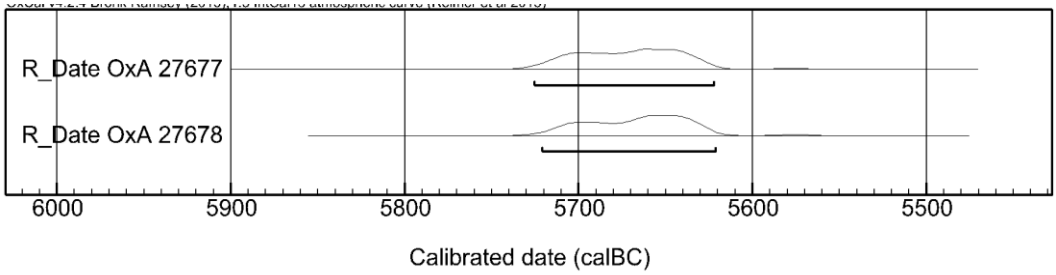

fig. S1. Calibrated age range of the two human specimens from Devil's Gate (OxCal version 4.2.4).

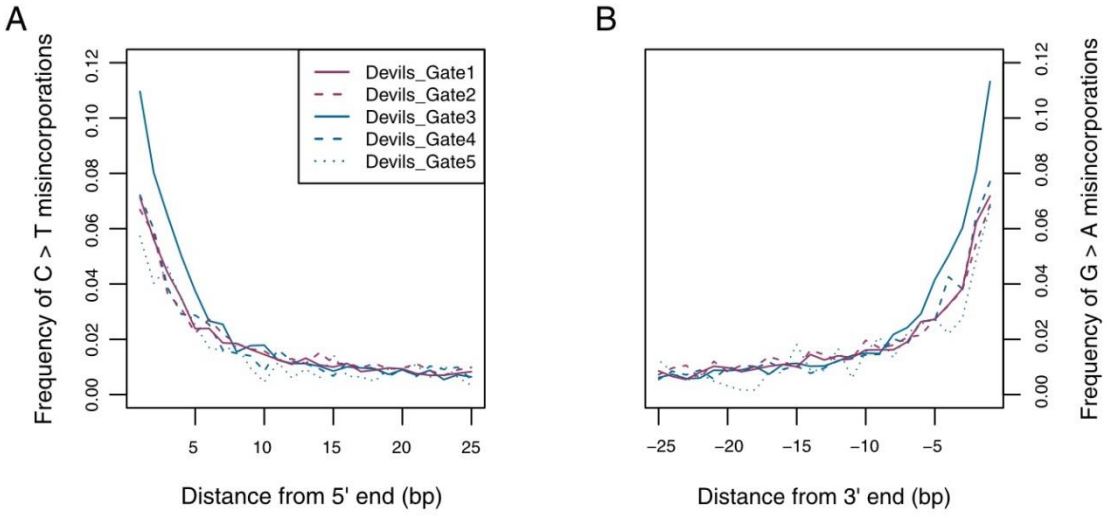

fig. S2. Damage patterns for samples from Devil's Gate. Plots show mismatch frequency relative to the reference genome as a function of read position. (**A**) shows the frequency of C to T misincorporations at the 5' ends of reads while (**B**) shows the frequency of G to A transitions at the 3' ends of reads.

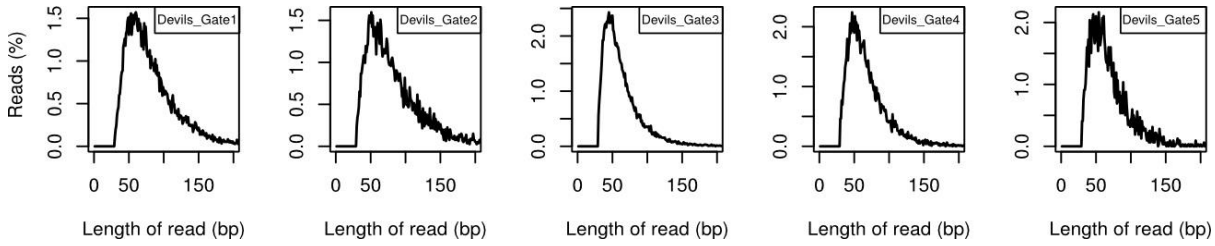

fig. S3. Sequence length distribution for samples from Devil's Gate. All samples have sequences in the range expected for ancient DNA.

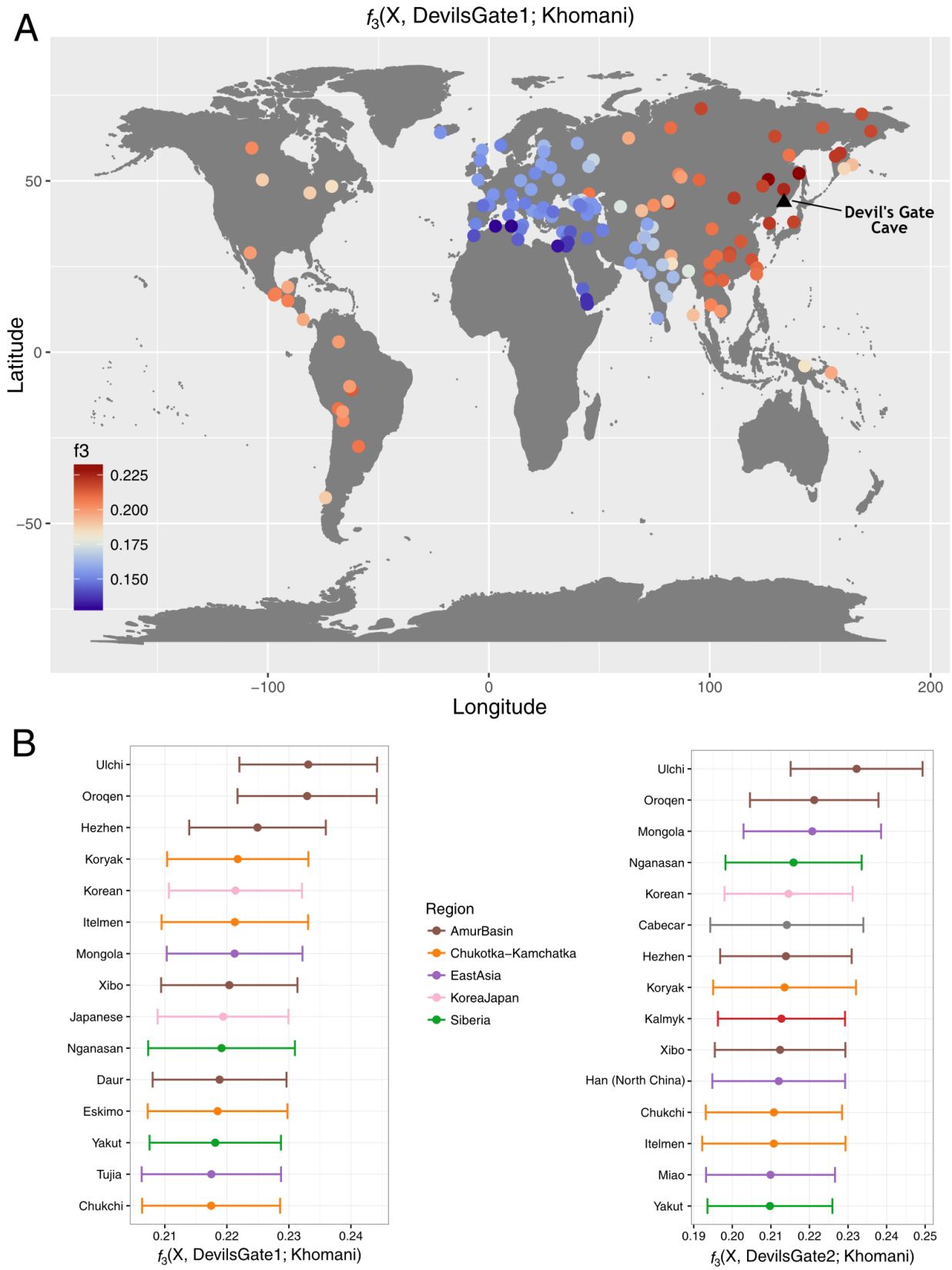

fig. S4. Outgroup f3 statistics on PMDtools-filtered data. Outgroup f_3 measuring shared drift between samples from Devil's Gate (black triangle shows sampling location) and modern populations with respect to an African outgroup (Khomani), using only reads with a PMD score of at least 3(9). (**A**) Map of the whole world. (**B**) 15 populations with the highest shared drift with Devil's Gate, color coded by regions as on Figure PCA. Error bars represent one standard error.

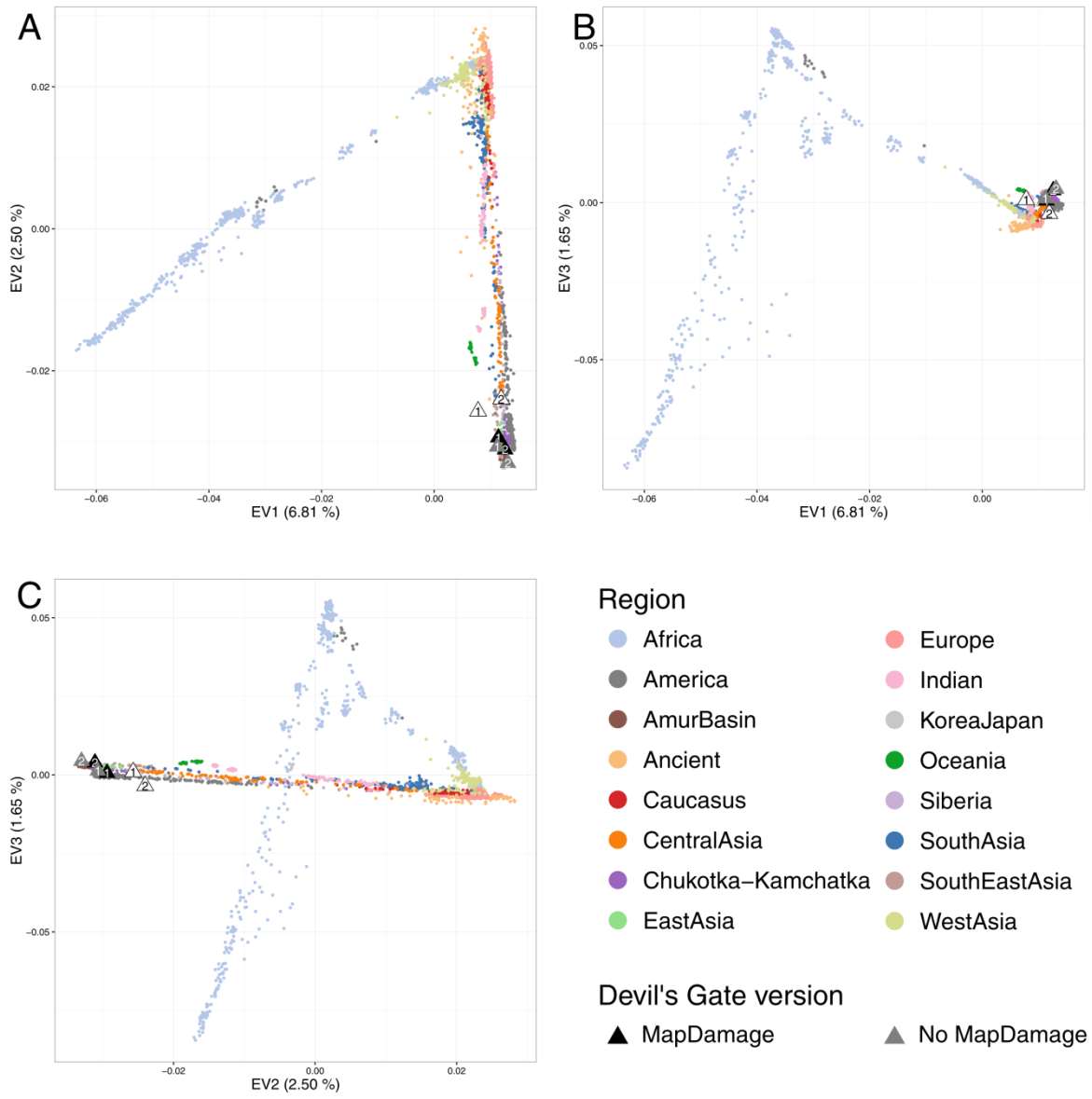

fig. S5. PCA on all SNPs using the worldwide panel. Black (MapDamage treated), gray (not MapDamage treated) and white (PMD-filtered, not MapDamage treated) symbols mark DevilsGate1 and DevilsGate2, projected upon the principal components as defined by the modern panel. Proportion of variance explained is displayed in parentheses on the axes. (**A**) Components 1 and 2. (**B**) Components 1 and 3. (**C**) Components 2 and 3.

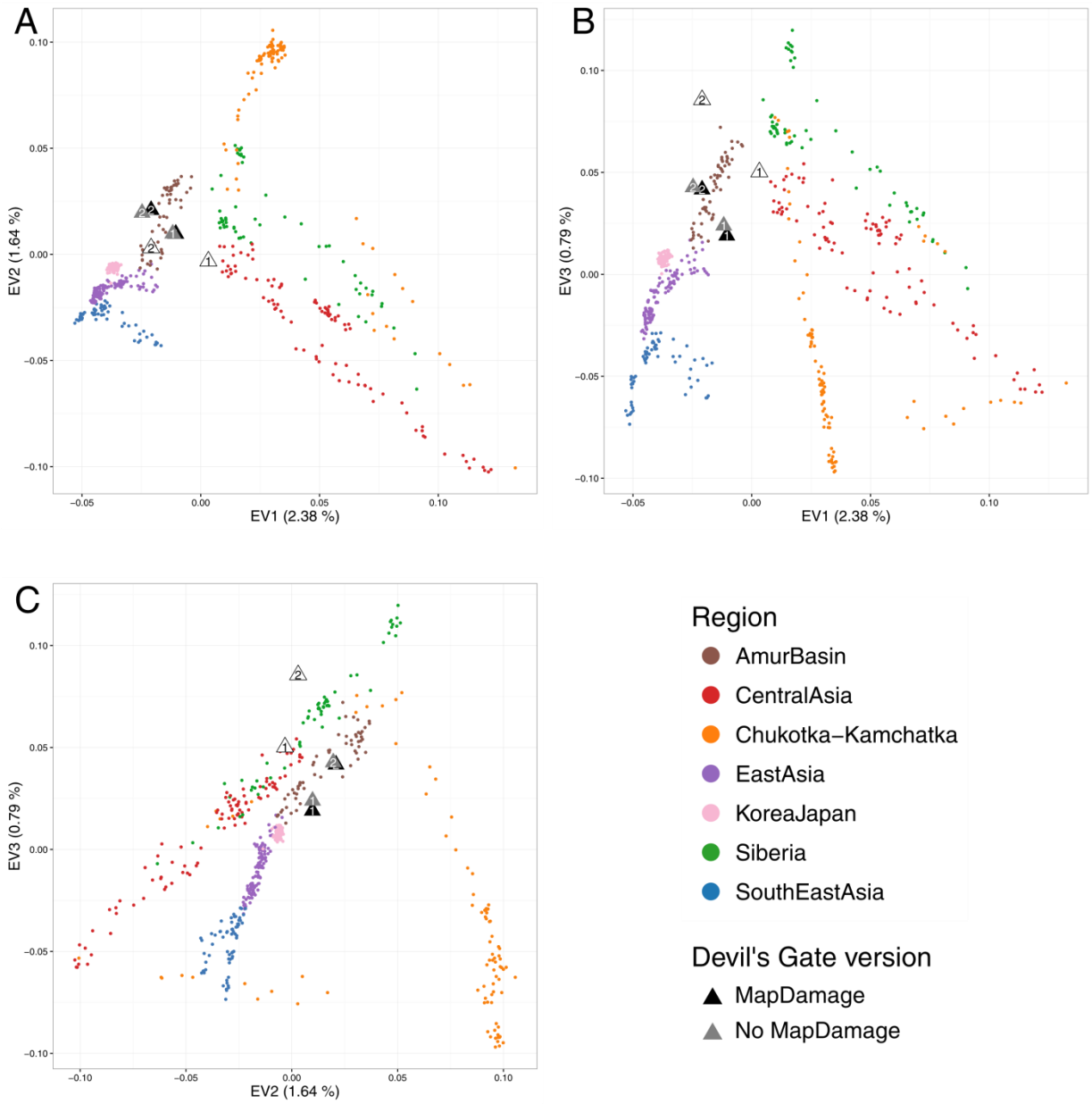

fig. S6. PCA on transversion SNPs using the worldwide panel. Black (MapDamage treated), gray (not MapDamage treated) and white (PMD-filtered, not MapDamage treated) symbols mark DevilsGate1 and DevilsGate2, projected upon the principal components as defined by the modern panel. Proportion of variance explained is displayed in parentheses on the axes. (**A**) Components 1 and 2. (**B**) Components 1 and 3. (**C**) Components 2 and 3.

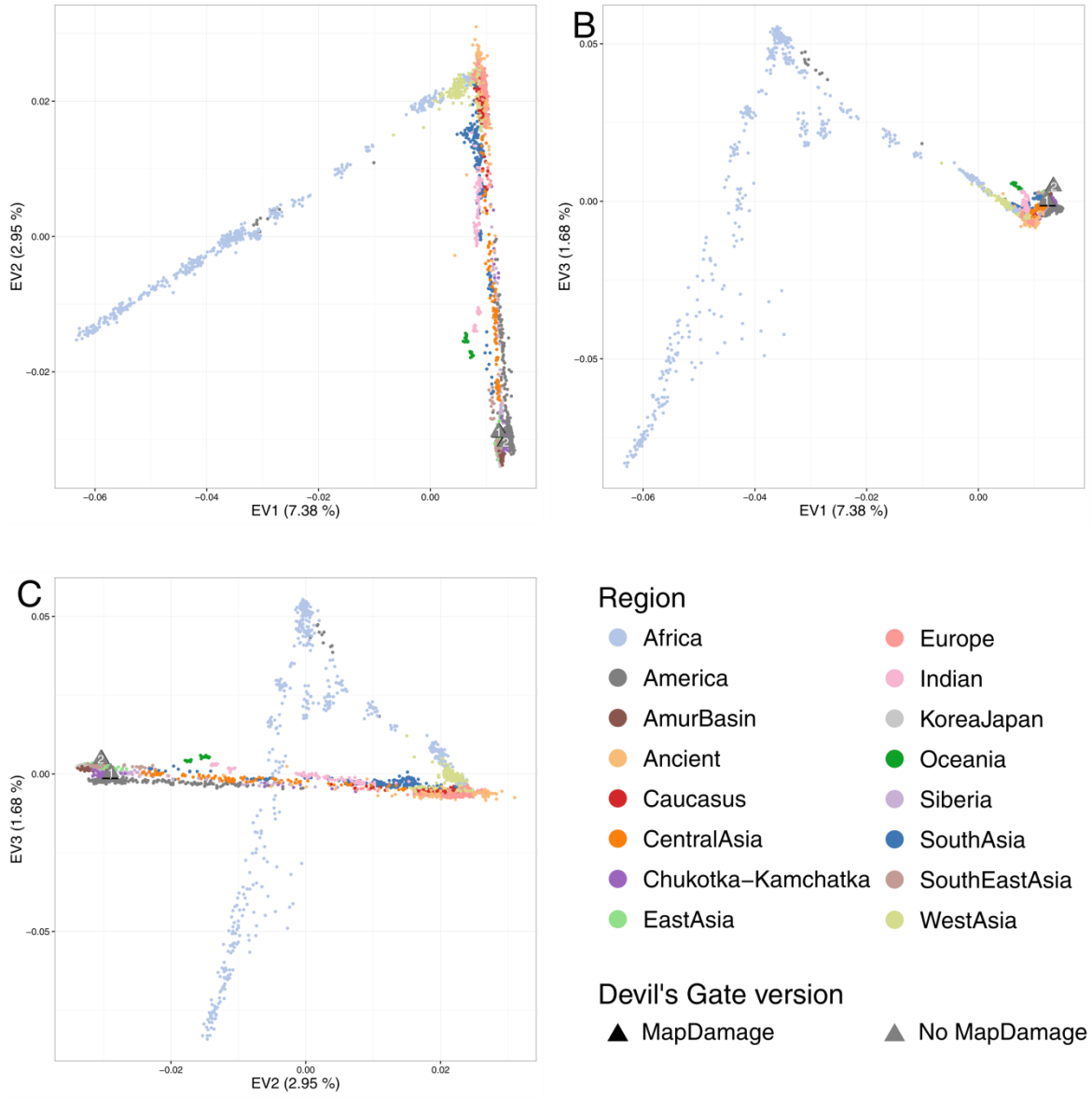

fig. S7. PCA on all SNPs using the regional panel. Black (MapDamage treated) and gray (not MapDamage treated) symbols mark DevilsGate1 and DevilsGate2, projected upon the principal components as defined by the modern panel. Proportion of variance explained is displayed in parentheses on the axes. (**A**) Components 1 and 2. (**B**) Components 1 and 3. (**C**) Components 2 and 3.

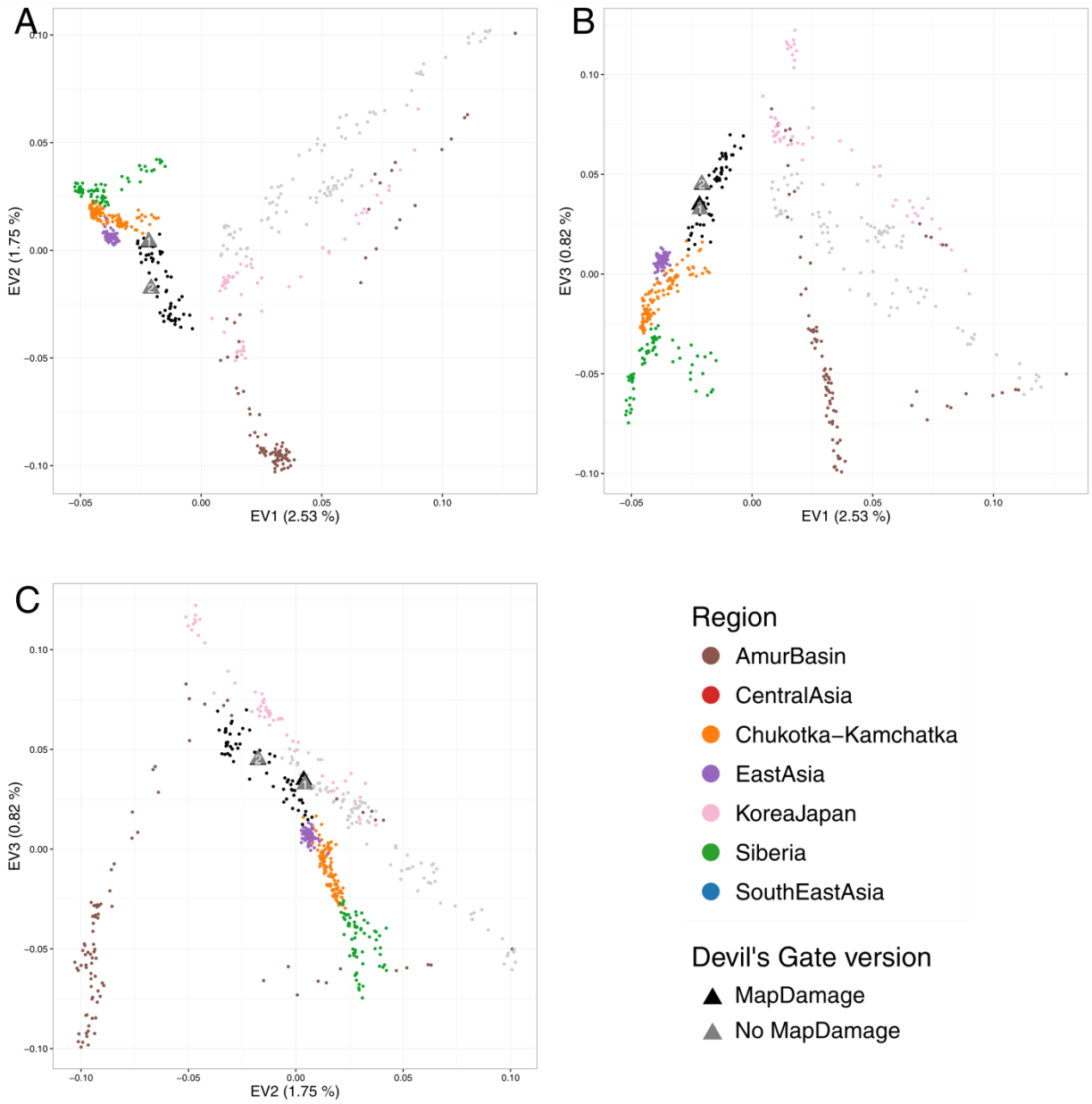

fig. S8. PCA on transversion SNPs using the regional panel. Black (MapDamage treated) and gray (not MapDamage treated) symbols mark DevilsGate1 and DevilsGate2, projected upon the principal components as defined by the modern panel. Proportion of variance explained is displayed in parentheses on the axes. (**A**) Components1 and 2. (**B**) Components 1 and 3. (**C**) Components 2 and 3.

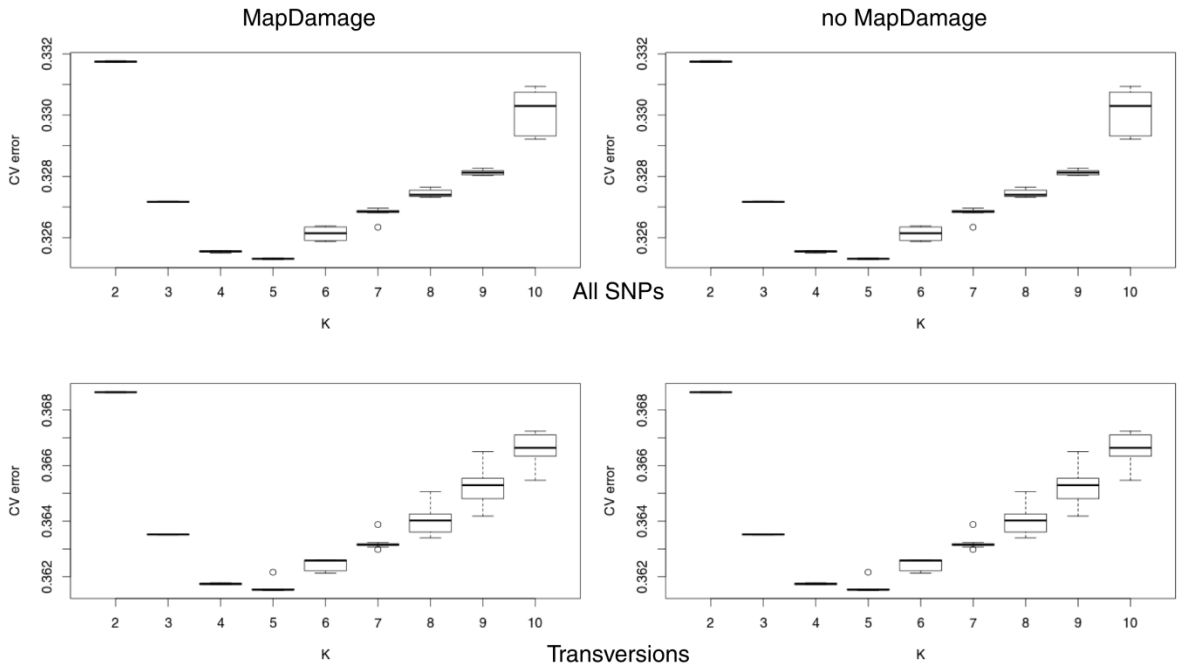

fig. S9. ADMIXTURE analysis cross-validation (CV) error as a function of the number of clusters (K) for the regional panel using all SNPs (top row) or transversions only (bottom row) and with (left column) or without (right column) MapDamage treatment. The lowest mean value was attained at K=5.

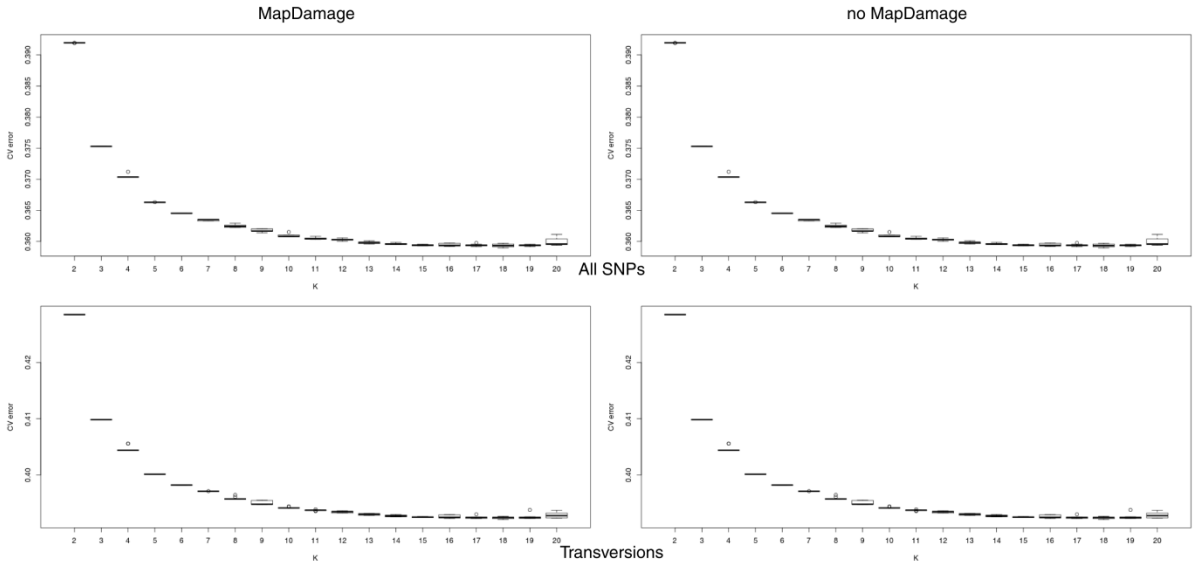

fig. S10. ADMIXTURE analysis CV error as a function of the number of clusters (K) for the world panel using all SNPs (top row) or transversions only (bottom row) and with (left column) or without (right column) MapDamage treatment. The lowest mean value was attained at K=18.

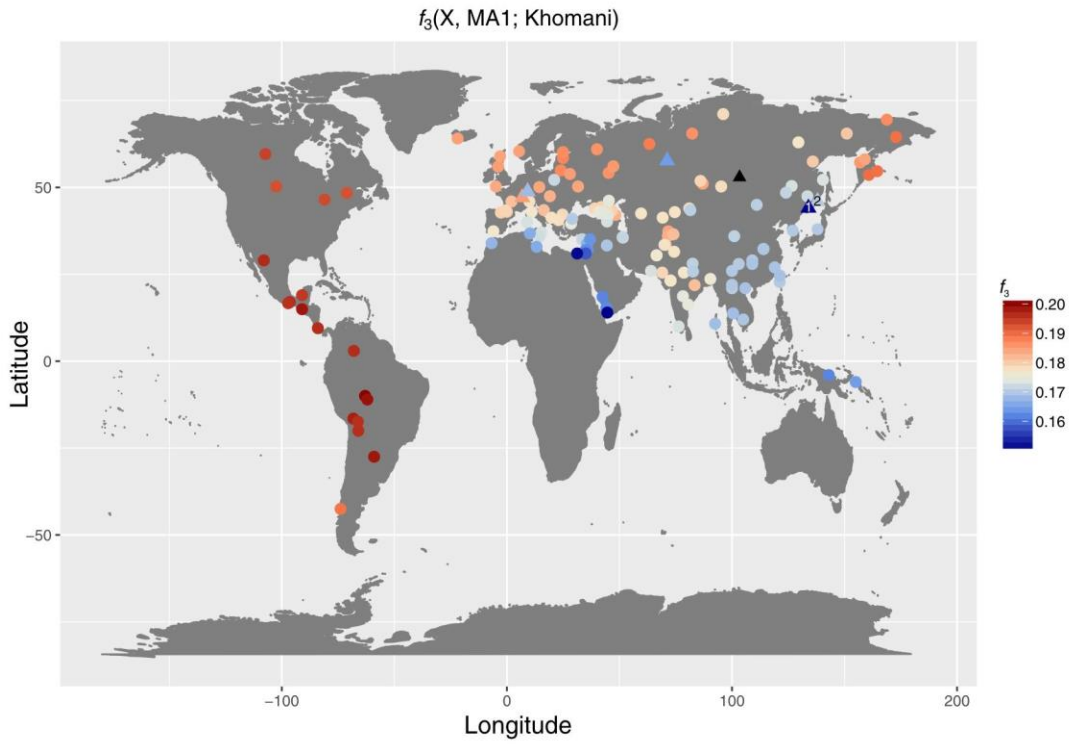

fig. S11. Outgroup f3 scores of the form f3(X, MA1; Khomani), with modern populations and selected ancient samples (DevilsGate1, DevilsGate2, Ust'-Ishim, Kotias, Loschbour, and Stuttgart), using all SNPs, with f3 > 0.15 displayed. Modern populations are displayed as circles and ancients as triangles. The black triangle marks the location of MA1.

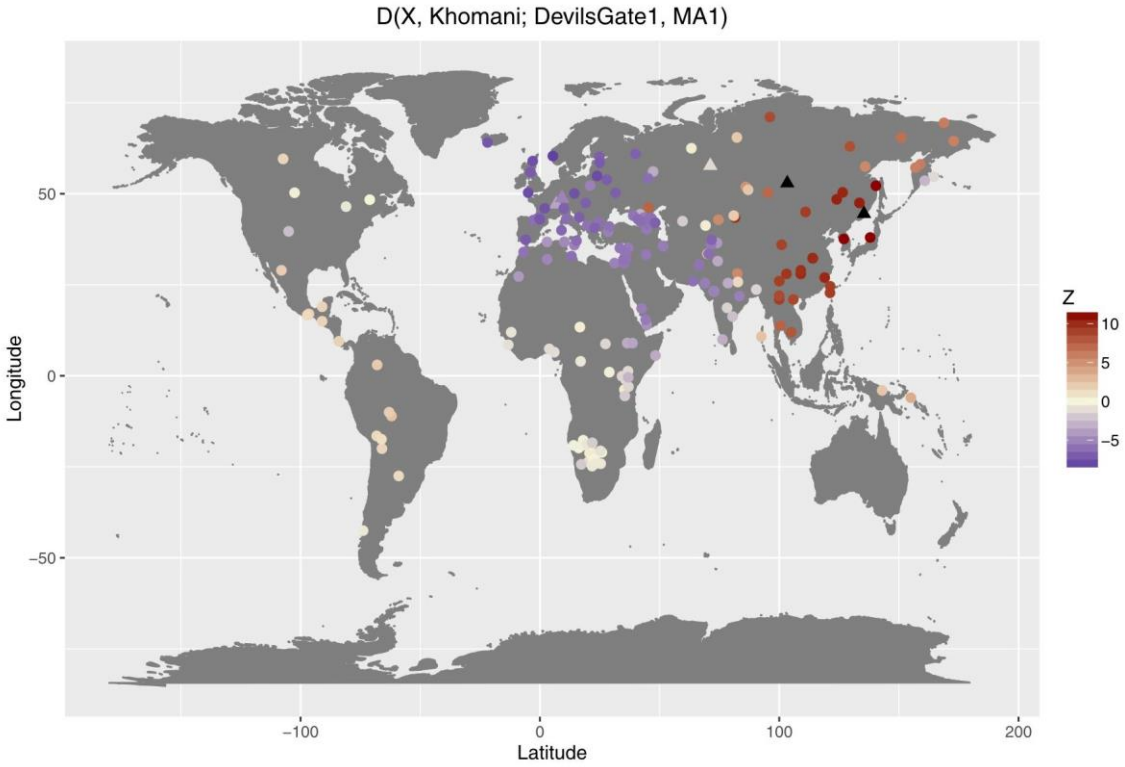

fig. S12. D scores of the form D(X, Khomani; MA1, DevilsGate1), with all modern populations in our panel and selected ancient samples, using all SNPs. Selected ancient samples (Ust'Ishim, Kotias, Loschbour and Stuttgart) are displayed as triangles. The black triangles mark the location of MA1 and Devil's Gate.

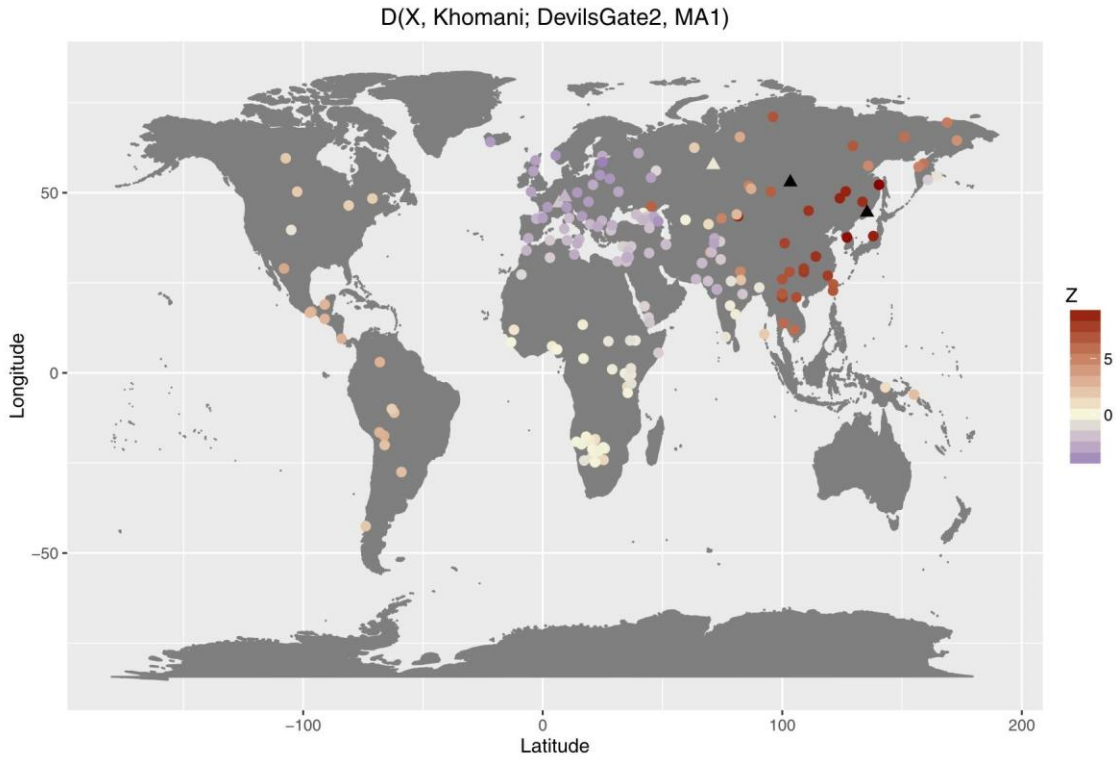

fig. S13. D scores of the form D(X, Khomani; MA1, DevilsGate1), with all modern populations in our panel and selected ancient samples, using all SNPs. (Ust'Ishim, Kotias, Loschbour and Stuttgart) displayed. The black triangles mark the location of MA1 and Devil's Gate.

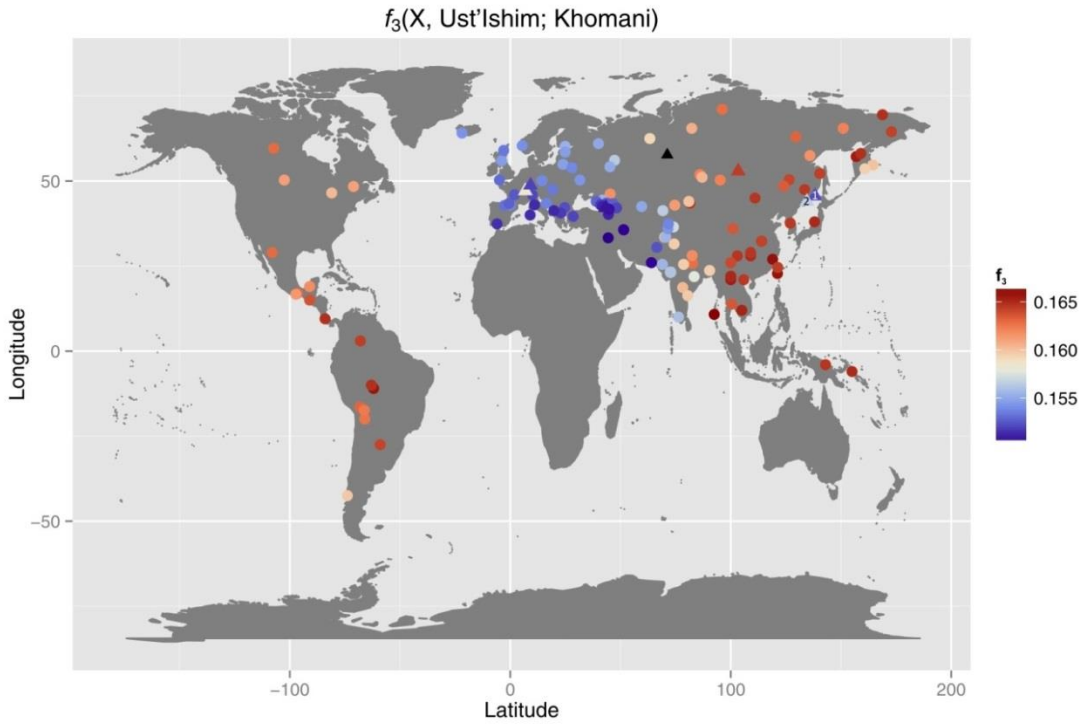

fig. S14. Outgroup f3 scores of the form f3(X, Ust'-Ishim; Khomani), with modern populations and selected ancient samples (MA1, Kotias, Loschbour, and Stuttgart), using all SNPs, with f3 > 0.15 displayed. The black triangle marks the location of Ust'Ishim.

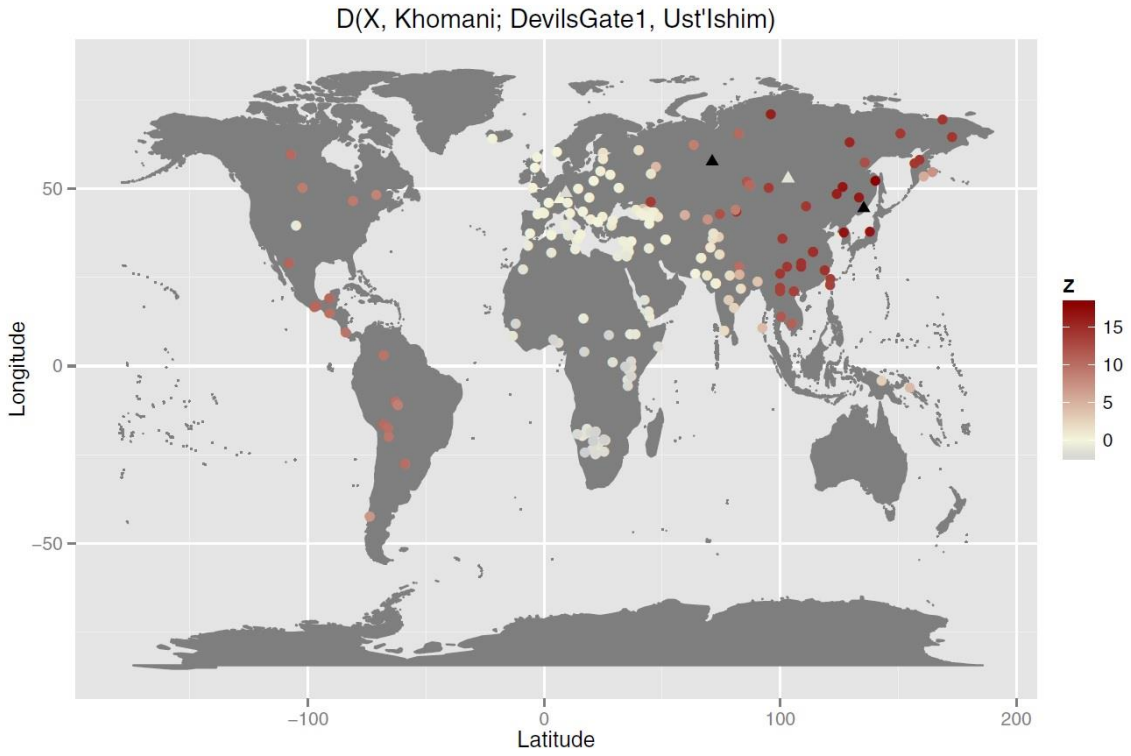

fig. S15. D scores of the form D(X, Khomani; Ust'-Ishim, DevilsGate1), with all modern populations in our panel and selected ancient samples, using all SNPs. (MA1, Kotias, Loschbour and Stuttgart) are displayed as triangles. The black triangles mark the location of MA1 and Devil's Gate.

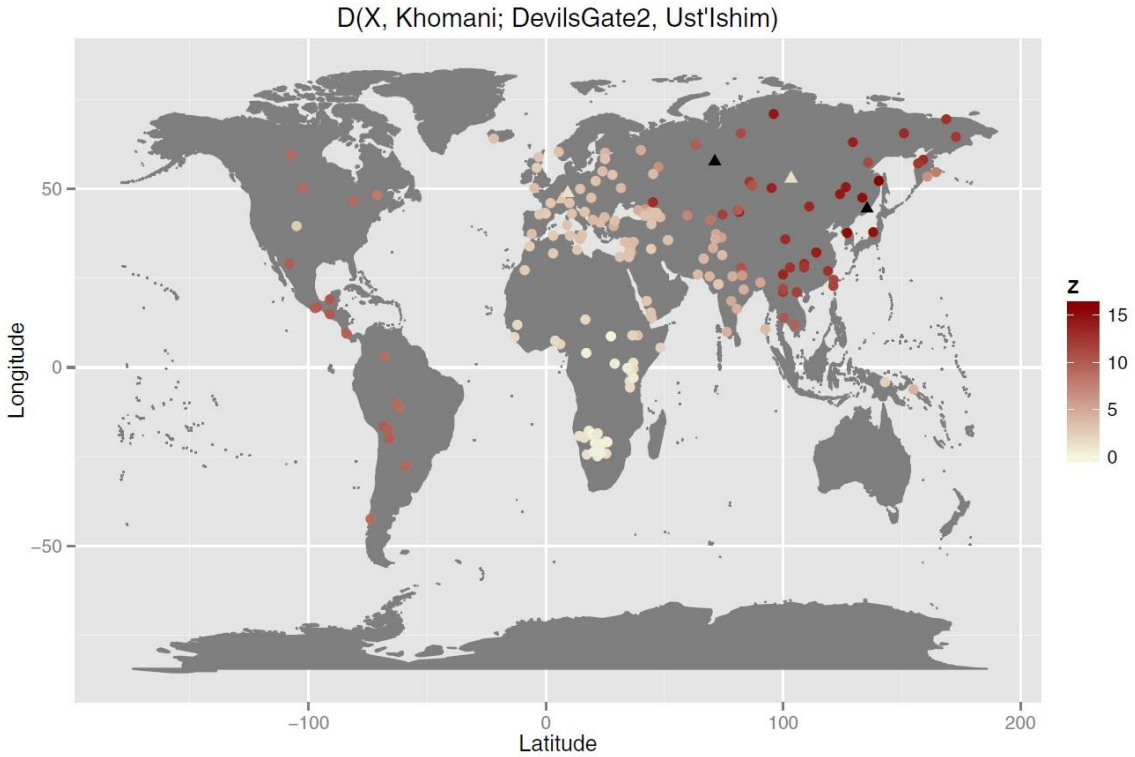

fig. S16. D scores of the form D(X, Khomani; Ust'-Ishim, DevilsGate2), with all modern populations in our panel and selected ancient samples, using all SNPs. Selected ancient samples(MA1, Kotias, Loschbour and Stuttgart) are displayed as triangles. The black triangles mark the location of MA1 and Devil's Gate.

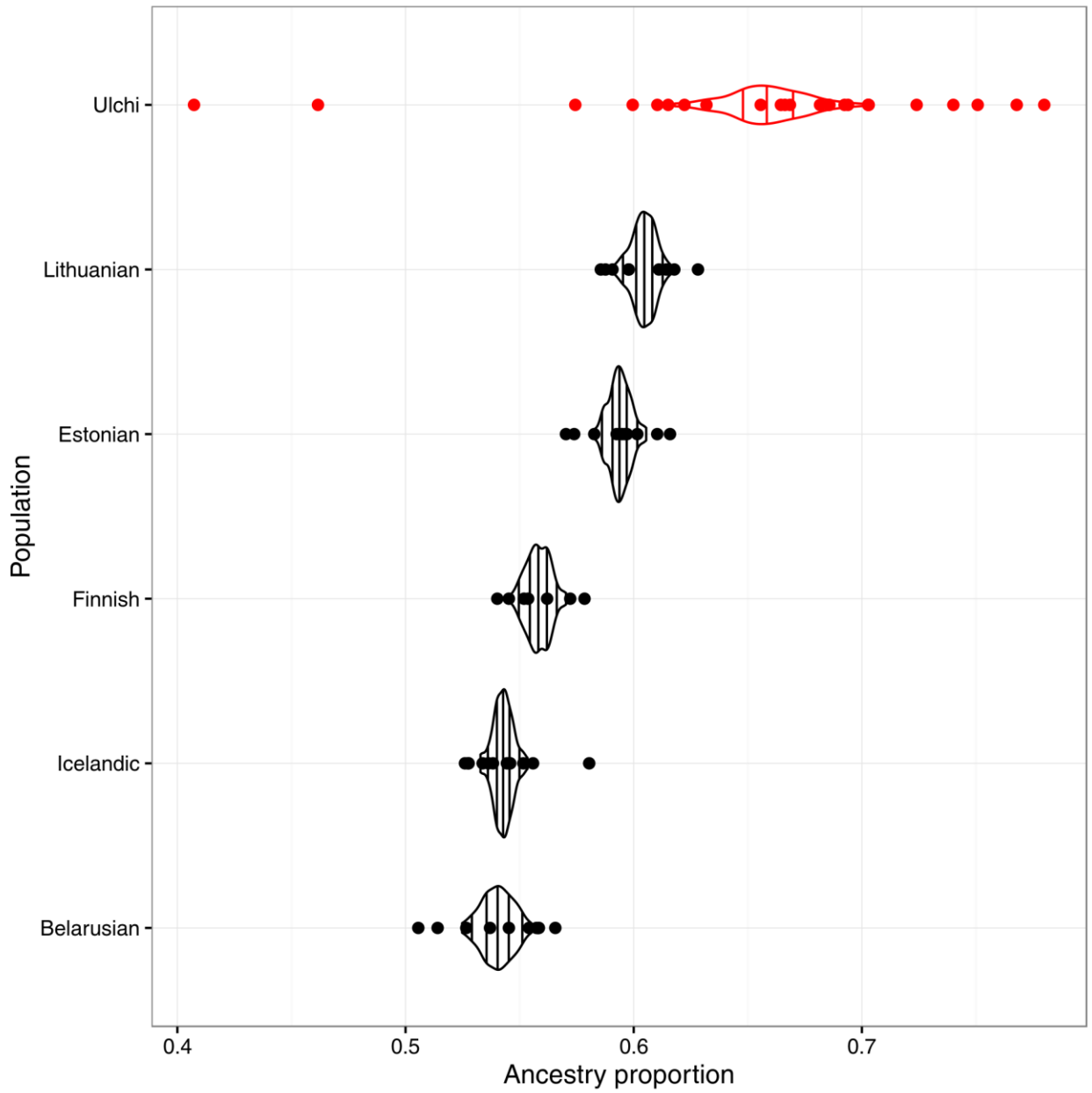

fig. S17. Comparison of Devil's Gate–related ancestry in the Ulchi and European hunter-gatherer–related ancestry in European populations. Dots mark ancestry proportions inferred by ADMIXTURE for each individual in the population (using all SNPs and MapDamage-treated versions of Devil's Gate), overlaid with violin plots of the distribution of population-level mean proportions from 100 bootstrap replicates on SNPs and individuals. Vertical lines in the violins mark the 5%, 25%, 50%, 75% and 95% quantiles. The five European populations with the highest mean proportion are shown.

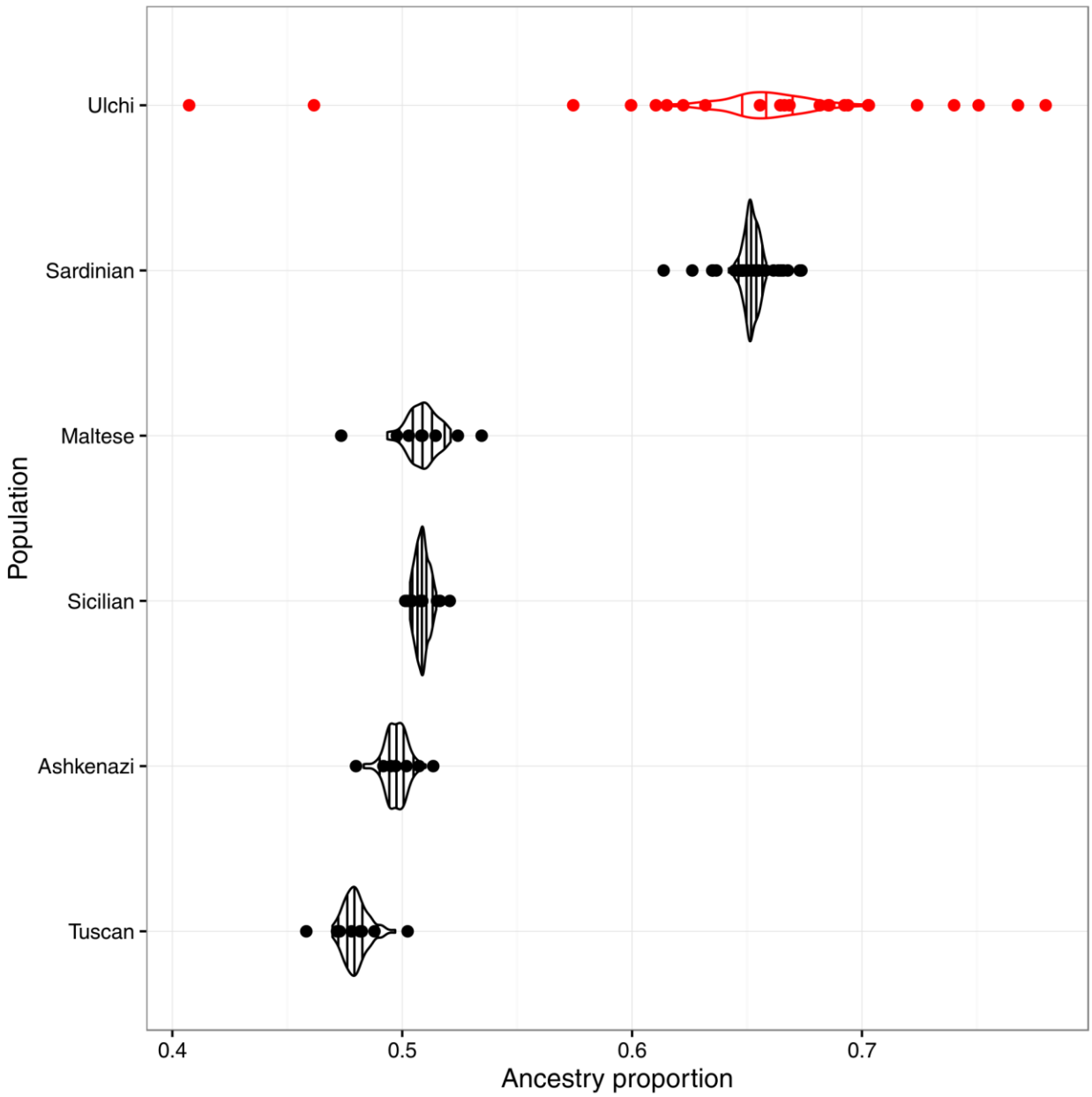

fig. S18. Comparison of Devil's Gate–related ancestry in the Ulchi and Early European farmer–related ancestry in European populations. Dots mark ancestry proportions inferred by ADMIXTURE for each individual in the population (using all SNPs and MapDamage-treated versions of Devil's Gate), overlaid with violin plots of the distribution of population-level mean proportions from 100 bootstrap replicates on SNPs and individuals. Vertical lines in the violins mark the 5%, 25%, 50%, 75% and 95% quantiles. The five European populations with the highest mean proportion are shown.

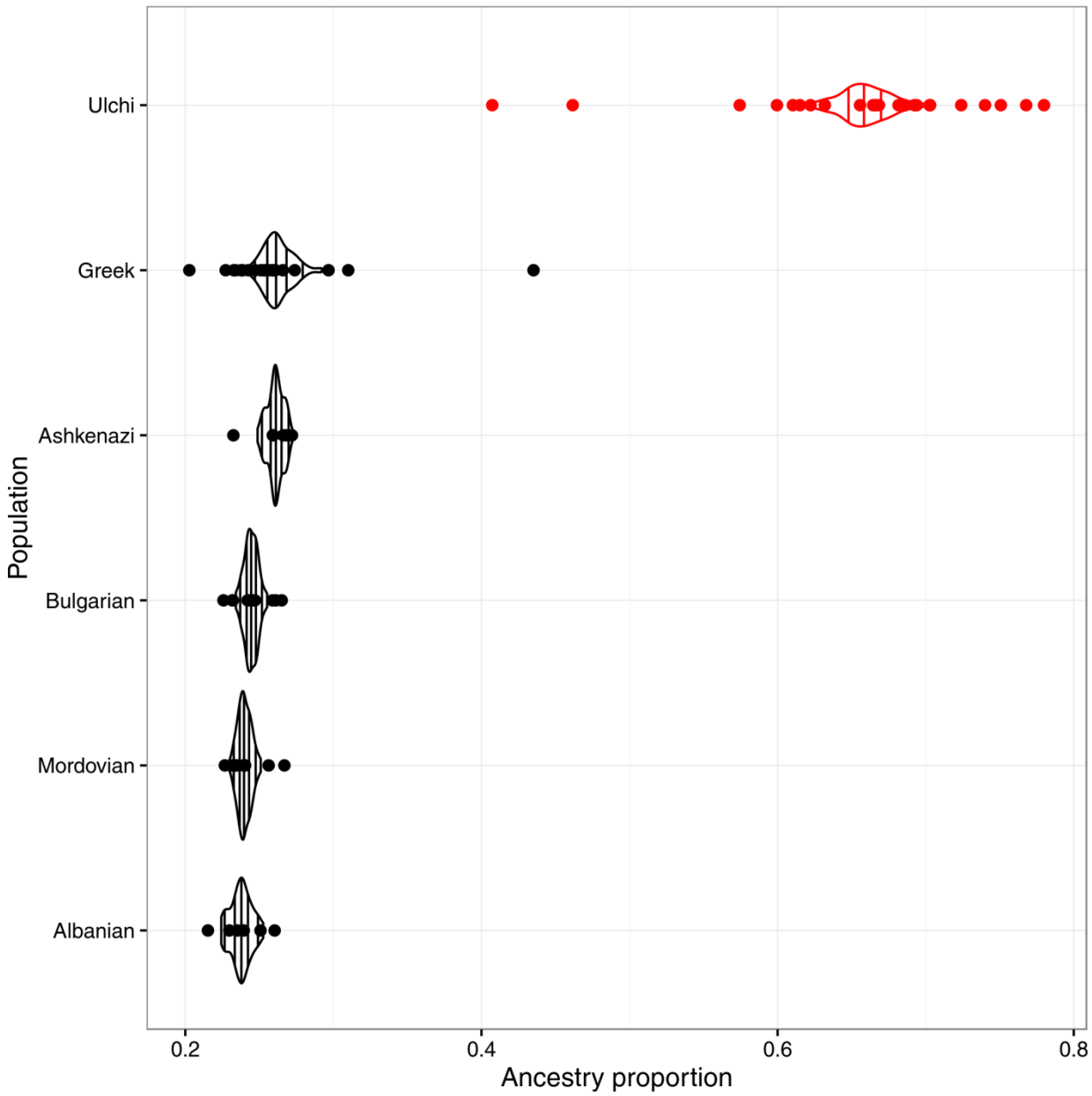

fig. S19. Comparison of Devil's Gate–related ancestry in the Ulchi and Bronze Age Steppe–related ancestry in European populations. Dots mark ancestry proportions inferred by ADMIXTURE for each individual in the population (using all SNPs and MapDamage-treated versions of Devil's Gate), overlaid with violin plots of the distribution of population-level mean proportions from 100 bootstrap replicates on SNPs and individuals. Vertical lines in the violins mark the 5%, 25%, 50%, 75% and 95% quantiles. The five European populations with the highest mean proportion are shown.

SUPPLEMENTARY TABLES

table S1. Details of sample preparation and sequencing. Extraction codes: 1*, DNA extracted from first lysis buffer; 2*, DNA extracted from second lysis buffer; bp, base pairs. High quality non-clonal reads refers to reads with mapping quality ⩾30 and of length ⩾30. Coverage refers to average genomic depth of coverage.

Sample	Library ID	Extraction	Skeletal element	Sequencing Instrument	Sequencing facility	Sequencing length and type	Reads	Aligned reads	%	High quality non-clonal reads	Coverage (x)
DevilsGate4	MOS2A.E1	1*	molar	Illumina MiSeq	TrinSeq, Dublin	150bp paired end	7,096,341	25,455	0.36	10792	0.0002
	MOS2C.E1	1*	cranial fragment	Illumina MiSeq	TrinSeq, Dublin	150bp paired end	433,003	1,510	0.35		
DevilsGate3	MOS3.2A.E1	1*	molar (root)	Illumina MiSeq	TrinSeq, Dublin	150bp paired end	535,857	14,895	2.78	31184	0.0006
	MOS3.3A.E1	1*	molar (crown)	Illumina MiSeq	TrinSeq, Dublin	150bp paired end	560,654	26,440	4.72		
DevilsGate2	MOS4A.E1	1*	molar	Illumina HiSeq	Beijing Genomics Institute, China	50bp single end	6,775,617	132,363	1.95	1,469,128	0.0333
	MOS4A.E1	1*	molar	Illumina HiSeq	Beijing Genomics Institute, China	50bp single end	108,566,893	2,003,142	1.85		
	MOS4A.E1	1*	molar	Illumina MiSeq	TrinSeq, Dublin	150bp paired end	618,270	11,689	1.89		
	MOS4A.E2	2*	molar	Illumina MiSeq	TrinSeq, Dublin	150bp paired end	338,589	7,172	2.12		
	MOS4.C3.E1	1*	cranial fragment	Illumina MiSeq	TrinSeq, Dublin	150bp paired end	532,719	2,822	0.53		
DevilsGate1	MOS5A.E1	1*	molar	Illumina HiSeq	Beijing Genomics Institute, China	50bp single end	3,775,905	127,260	3.37	2,625,622	0.0759
	MOS5A.E1	1*	molar	Illumina HiSeq	Beijing Genomics Institute, China	50bp single end	34,240,237	1,129,604	3.30		
	MOS5A.E1	1*	molar	Illumina MiSeq	TrinSeq, Dublin	150bp paired end	584,305	20,133	3.45		
	MOS5A.E2	2*	molar	Illumina MiSeq	TrinSeq, Dublin	150bp paired end	444,709	18,755	4.22		
	MOS5B.R1.E2	2*	molar (root)	Illumina MiSeq	TrinSeq, Dublin	70bp single end	2,632,825	216,661	8.23		
	MOS5B.R1.E2	2*	molar (root)	Illumina HiSeq	Theragen BiO Institute	100 bp paired end	60,490,976	2,043,091	3.38		
DevilsGate5	MOS6.C1.E1	1*	cranial fragment	Illumina MiSeq	TrinSeq, Dublin	150bp paired end	518,401	5,457	1.05	3,187	0.0001

table S2. mtDNA contamination estimates.

Sample	No. of positions	No. of consensus bases	No. of non-consensus bases	Proportion	95% C.I.
DevilsGate1	37	455	4	0.871%	0.280% - 2.373%
DevilsGate2	31	168	1	0.592%	0.030% - 3.753%

table S3. Admixture f3(Source1, Source2; Target) for the Ulchi with z < −1 using all SNPs. All pairs without a sample from Devil's Gate gave Z>-1, regardless of whether all SNPs were used or only those called in DevilsGate1 or DevilsGate2. Additional sample information is available from extended data table S1.

Source1	Source2	Target	f_3	SE	Z	SNPs
DevilsGate1MapDamage	MA1	Ulchi	-0.00647	0.00502	-1.287	15582
DevilsGate1	MA1	Ulchi	-0.00596	0.00506	-1.179	15582

table S4. Admixture f3(Source1, Source2; Target) for the Ulchi with z < −1 using only transversion SNPs. All pairs without a sample from Devil's Gate gave Z>-1, regardless of whether all SNPs were used or only those called in DevilsGate1 or DevilsGate2. Additional sample information is available from extended data table S1.

Source1	Source2	Target	f_3	SE	Z	SNPs
DevilsGate2	Motala12	Ulchi	-0.02081	0.01448	-1.438	1606
DevilsGate1MapDamage	Karelia_HG	Ulchi	-0.01955	0.01201	-1.628	2337
DevilsGate1	Karelia_HG	Ulchi	-0.01923	0.01191	-1.614	2337
DevilsGate1MapDamage	MA1	Ulchi	-0.01698	0.01075	-1.58	3004
DevilsGate1	MA1	Ulchi	-0.01462	0.01076	-1.359	3004
DevilsGate1MapDamage	Itelmen	Ulchi	-0.00533	0.00406	-1.313	4365
DevilsGate1MapDamage	Koryak	Ulchi	-0.00451	0.00355	-1.273	4368
DevilsGate1	Itelmen	Ulchi	-0.00517	0.00407	-1.27	4365

table S5. Admixture f3(Source1, Source2; Target) for the Sardinians using all SNPs and showing the 10 most significantly negative pairs. Additional sample information is available from extended data table S1.

Source1	Source2	Target	f_3	SE	Z	SNPs
Spain_EN	baArm	Sardinian	-0.028907	0.008386	-3.447	1984
Loschbour	baRem	Sardinian	-0.026789	0.00903	-2.967	2648
LaBrana	baRem	Sardinian	-0.03462	0.011897	-2.91	2512
Loschbour	Iraqi_Jew	Sardinian	-0.004893	0.001722	-2.842	9435
Spain_EN	Chuvash	Sardinian	-0.004701	0.001816	-2.588	5574
Xibo	Spain_EN	Sardinian	-0.006728	0.002612	-2.576	5571
Stuttgart	LaBrana	Sardinian	-0.012009	0.004766	-2.519	8627
Yemenite_Jew	Loschbour	Sardinian	-0.004339	0.001785	-2.43	9435
LBK_EN	irAltai	Sardinian	-0.007362	0.003034	-2.426	5214
Stuttgart	Karelia_HG	Sardinian	-0.014699	0.006069	-2.422	5212

table S6. Admixture f3(Source1, Source2; Target) for the Lithuanians using all SNPs and showing the 10 most significantly negative pairs. Additional sample information is available from extended data table S1.

Source1	Source2	Target	f_3	SE	Z	SNPs
Loschbour	Iraqi_Jew	Lithuanian	-0.014007	0.002063	-6.788	8617
Loschbour	Lezgin	Lithuanian	-0.012662	0.001965	-6.445	8629
Palestinian	Loschbour	Lithuanian	-0.010899	0.00172	-6.338	8660
Loschbour	Druze	Lithuanian	-0.010408	0.001667	-6.243	8653
Turkish	Loschbour	Lithuanian	-0.009566	0.001532	-6.243	8660
Loschbour	Armenian	Lithuanian	-0.011644	0.001879	-6.196	8634
Loschbour	Georgian_Jew	Lithuanian	-0.012235	0.00201	-6.088	8626
Loschbour	Georgian	Lithuanian	-0.011701	0.001983	-5.9	8632
Loschbour	Chechen	Lithuanian	-0.011107	0.00189	-5.875	8630
Loschbour	Adygei	Lithuanian	-0.010003	0.001713	-5.839	8637

EXTENDED DATA FIGURES

- **extended data fig. S1. Results from ADMIXTURE analysis using the regional panel, all SNPs, and MapDamage treatment on samples from Devil's Gate and setting the number of clusters to *K* = 2 to 10.** Minimal cross-validation error was attained at K=5.

- **extended data fig. S2. Results from ADMIXTURE analysis using the regional panel, transversion SNPs, and MapDamage treatment on samples from Devil's Gate and setting the number of clusters to *K* = 2 to 10.** Minimal cross-validation error was attained at K=5.

- **extended data fig. S3. Results from ADMIXTURE analysis using the regional panel, all SNPs, and no MapDamage treatment on samples from Devil's Gate and setting the number of clusters to *K* = 2 to 10.** Minimal cross-validation error was attained at K=5.

- **extended data fig. S4. Results from ADMIXTURE analysis using the regional panel, transversion SNPs, and no MapDamage treatment on samples from Devil's Gate and setting the number of clusters to *K* = 2 to 10.** Minimal cross-validation error was attained at K=5.

- **extended data fig. S5. Results from ADMIXTURE analysis using the total panel, all SNPs, and MapDamage treatment on samples from Devil's Gate and setting the number of clusters to *K* = 2 to 20.** Minimal cross-validation error was attained at K=5.

- **extended data fig. S6. Results from ADMIXTURE analysis using the total panel, transversion SNPs, and MapDamage treatment on samples from Devil's Gate and setting the number of clusters to *K* = 2 to 20.** Minimal cross-validation error was attained at K=5.

- **extended data fig. S7. Results from ADMIXTURE analysis using the total panel, all SNPs, and no MapDamage treatment on samples from Devil's Gate and setting the number of clusters to *K* = 2 to 20.** Minimal cross-validation error was attained at K=5.

- **extended data fig. S8. Results from ADMIXTURE analysis using the total panel, transversion SNPs, and no MapDamage treatment on samples from Devil's Gate and setting the number of clusters to *K* = 2 to 20.** Minimal cross-validation error was attained at K=5.

EXTENDED DATA TABLES

- **extended data table S1. Sample information.** Additional information about the modern and ancient populations in our panel, including groupings used throughout the manuscript.

- **extended data table S2. ADMIXTURE proportions.** Proportions of the three ancestral components (European Hunter-Gatherer, Early European Farmer, Bronze Age Steppe) in modern European populations, with the five populations with the highest population-level mean proportions shown, as reported by ADMIXTURE runs from the regional panel at k=8 for the Ulchi and the global panel at k=18 for modern European populations. Results using all SNPs and transversions only, as well as MapDamage- or non-MapDamage-treated Devil's Gate versions are shown.

- **extended data table S3. Outgroup f_3 statistics for Devil's Gate.** f_3 statistics of the form f_3(DevilsGate, X; Khomani) for all populations X in our panel compared to DevilsGate1 and DevilsGate2, using all SNPs or transversions only.

- **extended data table S4. Outgroup f_3 and space.** Relationship between outgroup f_3 (Devil's Gate; Khomani) and distance on land from Devil's Gate. The highest distance considered was chosen to acquire the highest correlation in steps of 500 km.

- **extended data table S5. Outgroup f_3 for MA1 and Ust'-Ishim.** f_3 statistics of the form f_3 (Ancient, X; Khomani) for all populations X in our panel, where Ancient is either MA1 or Ust'Ishim, using all SNPs or transversions only.

- **extended data table S6. *D* scores for MA1 and Ust'-Ishim.** D statistics of the form D(Ancient, Devil's Gate; X, Khomani) for all populations X in our panel, where Ancient is either MA1 or Ust'Ishim, using all SNPs or transversions only.

- **extended data table S7. *D* scores for the Ulchi.** D statistics of the form D(Ulchi, Devil's Gate; X, Khomani) for all populations X in our panel, using all SNPs or transversions only.

- **extended data table S8. Admixture f_3 for the Koreans and the Japanese.** f_3 statistics of the form f_3(X, Y; Target) for all populations X in our panel, where Target is either the Koreans or the Japanese, using all SNPs, transversions only, SNPs called in DevilsGate1 or SNPs called in DevilsGate2.

- **extended data table S9. Phenotypes of interest.** Results of imputed SNPs with known biological function.