



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

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

Suggested citation referring to the original publication:
Science Advances 3 (2017) 2, Art. e1601877
DOI <https://doi.org/10.1126/sciadv.1601877>
ISSN (online) 2375-2548

Postprint archived at the Institutional Repository of the Potsdam University in:
Postprints der Universität Potsdam
Mathematisch-Naturwissenschaftliche Reihe ; 791
ISSN 1866-8372
<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-439977>
DOI <https://doi.org/10.25932/publishup-43997>

EVOLUTIONARY GENETICS

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

Veronika Siska,^{1*} Eppie Ruth Jones,^{1,2} Sungwon Jeon,³ Youngjune Bhak,³ Hak-Min Kim,³ Yun Sung Cho,³ Hyunho Kim,⁴ Kyusang Lee,⁵ Elizaveta Veselovskaya,⁶ Tatiana Balueva,⁶ Marcos Gallego-Llorente,¹ Michael Hofreiter,⁷ Daniel G. Bradley,² Anders Eriksson,¹ Ron Pinhasi,^{8*†} Jong Bhak,^{3,4*†‡} Andrea Manica^{1*†}

2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

Ancient genomes have revolutionized our understanding of Holocene prehistory and, particularly, the Neolithic transition in western Eurasia. In contrast, East Asia has so far received little attention, despite representing a core region at which the Neolithic transition took place independently ~3 millennia after its onset in the Near East. We report genome-wide data from two hunter-gatherers from Devil's Gate, an early Neolithic cave site (dated to ~7.7 thousand years ago) located in East Asia, on the border between Russia and Korea. Both of these individuals are genetically most similar to geographically close modern populations from the Amur Basin, all speaking Tungusic languages, and, in particular, to the Ulchi. The similarity to nearby modern populations and the low levels of additional genetic material in the Ulchi imply a high level of genetic continuity in this region during the Holocene, a pattern that markedly contrasts with that reported for Europe.

INTRODUCTION

Ancient genomes from western Asia have revealed a degree of genetic continuity between preagricultural hunter-gatherers and early farmers 12 to 8 thousand years ago (ka) (1, 2). In contrast, studies on southeast and central Europe indicate a major population replacement of Mesolithic hunter-gatherers by Neolithic farmers of a Near Eastern origin during the period 8.5 to 7 ka. This is then followed by a progressive "resurgence" of local hunter-gatherer lineages in some regions during the Middle/Late Neolithic and Eneolithic periods and a major contribution from the Asian Steppe later, ~5.5 ka, coinciding with the advent of the Bronze Age (3–5). Compared to western Eurasia, for which hundreds of partial ancient genomes have already been sequenced, East Asia has been largely neglected by ancient DNA studies to date, with the exception of the Siberian Arctic belt, which has received attention in the context of the colonization of the Americas (6, 7). However, East Asia represents an extremely interesting region as the shift to reliance on agriculture appears to have taken a different course from that in western Eurasia. In the latter region, pottery, farming, and animal husbandry were closely associated. In contrast, Early Neolithic societies in the Russian Far East, Japan, and Korea started to manufacture and use pottery and basketry 10.5 to 15 ka, but domesticated crops and livestock arrived several millennia later (8, 9). Because of the current lack of ancient genomes from East Asia, we do not know the extent to which this gradual Neolithic transition, which happened independently from the one taking place in western Eurasia, reflected actual

migrations, as found in Europe, or the cultural diffusion associated with population continuity.

RESULTS

Samples, sequencing, and authenticity

To fill this gap in our knowledge about the Neolithic in East Asia, we sequenced to low coverage the genomes of five early Neolithic burials (DevilsGate1, 0.059-fold coverage; DevilsGate2, 0.023-fold coverage; and DevilsGate3, DevilsGate4, and DevilsGate5, <0.001-fold coverage) from a single occupational phase at Devil's Gate (Chertovy Vorota) Cave in the Primorye Region, Russian Far East, close to the border with China and North Korea (see the Supplementary Materials). This site dates back to 9.4 to 7.2 ka, with the human remains dating to ~7.7 ka, and it includes some of the world's earliest evidence of ancient textiles (10). The people inhabiting Devil's Gate were hunter-fisher-gatherers with no evidence of farming; the fibers of wild plants were the main raw material for textile production (10). We focus our analysis on the two samples with the highest sequencing coverage, DevilsGate1 and DevilsGate2, both of which were female. The mitochondrial genome of the individual with higher coverage (DevilsGate1) could be assigned to haplogroup D4; this haplogroup is found in present-day populations in East Asia (11) and has also been found in Jomon skeletons in northern Japan (2). For the other individual (DevilsGate2), only membership to the M branch (to which D4 belongs) could be established. Contamination, estimated from the number of discordant calls in the mitochondrial DNA (mtDNA) sequence, was low {0.87% [95% confidence interval (CI), 0.28 to 2.37%] and 0.59% (95% CI, 0.03 to 3.753%)} on nonconsensus bases at haplogroup-defining positions for DevilsGate1 and DevilsGate2, respectively. Using schmutzi (12) on the higher-coverage genome, DevilsGate1 also gives low contamination levels [1% (95% CI, 0 to 2%); see the Supplementary Materials]. As a further check against the possible confounding effect of contamination, we made sure that our most important analyses [outgroup f_3 scores and principal components analysis (PCA)] were qualitatively replicated using only reads showing evidence of postmortem damage (PMD score of at least 3) (13), although these latter results had a high level of noise due to the low coverage (0.005X for DevilsGate1 and 0.001X for DevilsGate2).

¹Department of Zoology, University of Cambridge, Downing Street, Cambridge CB23EJ, U.K. ²Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. ³The Genomics Institute, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea. ⁴Geromics, Ulsan 44919, Republic of Korea. ⁵Clinomics Inc., Ulsan 4919, Republic of Korea. ⁶Institute of Ethnology and Anthropology, Russian Academy of Sciences, Moscow, Russia. ⁷Institute for Biochemistry and Biology, Faculty for Mathematics and Natural Sciences, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam-Golm, Germany. ⁸School of Archaeology and Earth Institute, University College Dublin, Dublin, Ireland.

*Corresponding author. Email: vs389@cam.ac.uk (V.S.); ron.pinhasi@ucd.ie (R.P.); jongbhak@genomics.org (J.B.); am315@cam.ac.uk (A.M.)

†These authors contributed equally to this work.

‡Adjunct professor at Seoul National University, Seoul, Republic of Korea.

Relation to modern populations

We compared the individuals from Devil's Gate to a large panel of modern-day Eurasians and to published ancient genomes (Fig. 1A) (4, 5, 14–17). On the basis of PCA (18) and an unsupervised clustering approach, ADMIXTURE (19), both individuals fall within the range of modern variability found in populations from the Amur Basin, the geographic region where Devil's Gate is located (Fig. 1), and which is today inhabited by speakers from a single language family (Tungusic). This result contrasts with observations in western Eurasia, where, because of a number of major intervening migration waves, hunter-gatherers of a similar age fall outside modern genetic variation (3, 20). We further confirmed the affinity between Devil's Gate and modern-day Amur Basin populations by using outgroup f_3 statistics in the form $f_3(\text{African}; \text{DevilsGate}, X)$, which measures the amount of shared genetic drift between a Devil's Gate individual and X , a modern or ancient population, since they diverged from an African outgroup. Modern populations that live in the same geographic region as Devil's Gate have the highest genetic affinity to our ancient genomes (Fig. 2), with a progressive decline in affinity with increasing geographic distance ($r^2 = 0.756$, $F_{1,96} = 301$, $P < 0.001$; Fig. 3), in agreement with neutral drift leading to a simple isolation-by-distance pattern. The Ulchi, traditionally fishermen who live geographically very close to Devil's Gate and are the only Tungusic-speaking population from the Amur Basin sampled in Russia (all other Tungusic speakers in our panel are from China), are genetically the most similar population in our panel. Other populations that show high affinity to Devil's Gate are the Oroqen and the Hezhen—both of whom, like the Ulchi, are Tungusic speakers from the Amur Basin—as well as modern Koreans and Japanese. Given their geographic distance from Devil's Gate (Fig. 3), Amerindian populations are unusually genetically close to samples from this site, in agreement with their previously reported relationship to Siberian and other north Asian populations (7).

Relation to ancient genomes from Asia

No previously published ancient genome shows marked genetic affinity to Devil's Gate: The top 50 populations in our outgroup f_3 statistic were all modern, an expected result given that all other ancient genomes are either geographically or temporally very distant from Devil's Gate. Among these ancient genomes, the closest to Devil's Gate are those from Steppe populations dating from the Bronze Age onward and Mesolithic hunter-gatherers from Europe, but these genomes are no closer to the Devil's Gate genomes than to genomes of modern populations from the same regions (for example, Tuvian, Kalmyk, Russian, or Finnish). The two ancient genomes geographically closest to Devil's Gate, Ust'-Ishim (~45 ka) and Mal'ta (MA1, 24 ka), also do not show high genetic affinity, probably because they both date to a much earlier time period. Of the two, MA1 is genetically closer to Devil's Gate, but it is equally as distant from Devil's Gate as it is from all other East Asians (figs. S14 to S16). A similar pattern is found for Ust'-Ishim, which is equally as distant to all Asians, including Devil's Gate; this is consistent with its basal position in a genealogical tree (figs. S17 to S19).

Continuity between Devil's Gate and the Ulchi

Because Devil's Gate falls within the range of modern human genetic variability in the Amur Basin in a number of analyses and shows a high genetic affinity to the Ulchi, we investigated the extent of genetic continuity in this region. To look for signals of additional genetic material in the Ulchi, we modeled them as a mixture of Devil's Gate

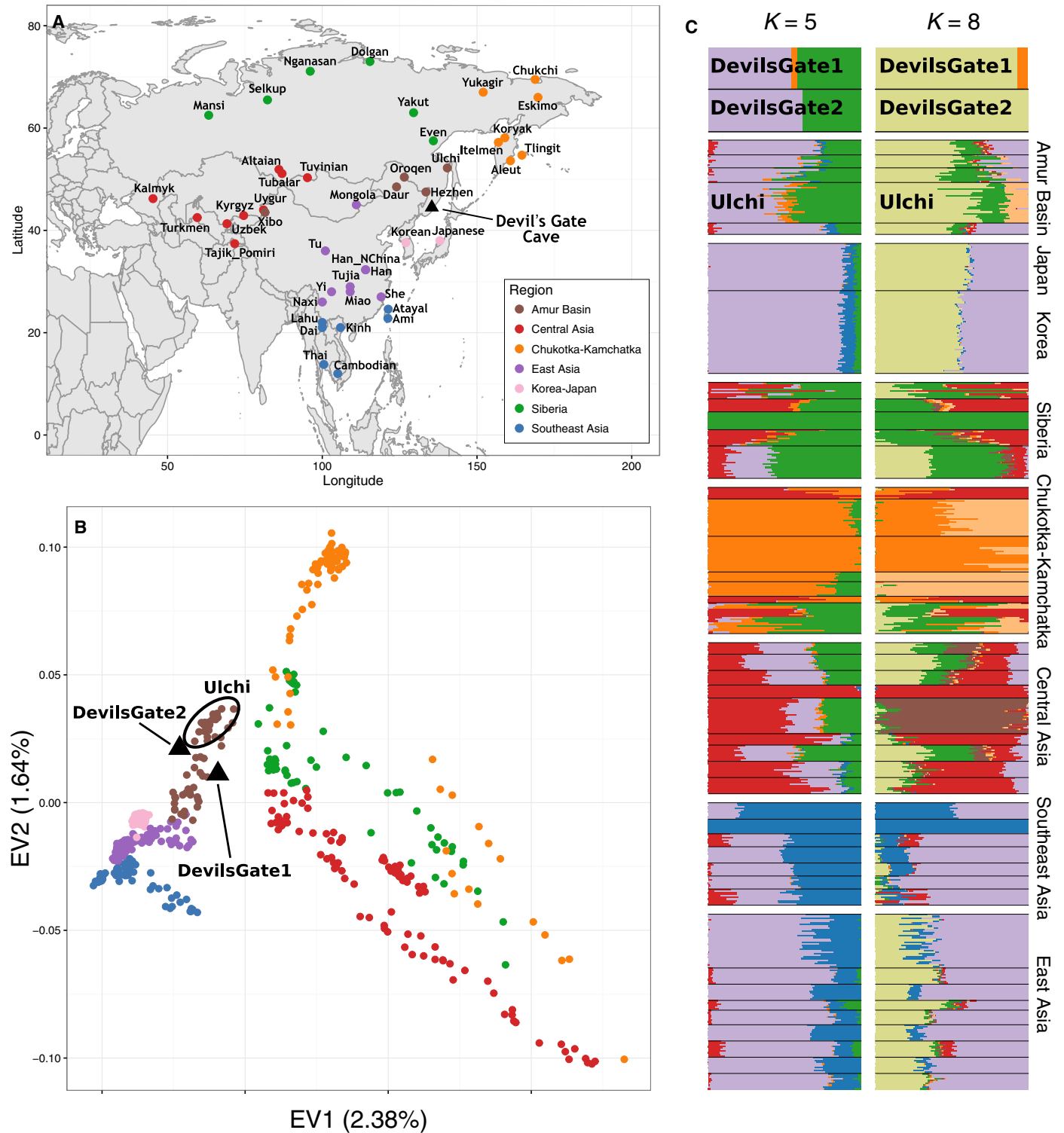
and other modern populations using admixture f_3 statistics. Despite a large panel of possible modern sources, the Ulchi are best represented by Devil's Gate alone without any further contribution (no admixture f_3 gave a significant negative result; tables S3 and S4). Because admixture f_3 can be affected by demographic events such as bottlenecks, we also tested whether Devil's Gate formed a clade with the Ulchi using a D statistic in the form $D(\text{African outgroup}, X; \text{Ulchi}, \text{Devil's Gate})$. A number of primarily modern populations worldwide gave significantly nonzero results ($|z| > 2$), which, together with the additional components for the Ulchi in the ADMIXTURE analysis, suggests that the continuity is not absolute. However, it should be noted that the higher error rates in the Devil's Gate sequence resulting from DNA degradation and low coverage can also decrease the inferred level of continuity. To compare the level of continuity between the Ulchi and the inhabitants of Devil's Gate to that between modern Europeans and European hunter-gatherers, we compared their ancestry proportions as inferred by ADMIXTURE. We found that the proportion of Devil's Gate-related ancestry in the Ulchi was significantly higher than the local hunter-gatherer-related ancestry in any European population ($P < 0.01$ from 100 bootstrap replicates for the five European populations with the highest mean hunter-gatherer-related component).

These results suggest a relatively high degree of continuity in this region; the Ulchi are likely descendants of Devil's Gate (or a population genetically very close to it), but the geographic and genetic connectivity among populations in the region means that this modern population also shows increased association with related modern populations. Compared to Europe, these results suggest a higher level of genetic continuity in northern East Asia over the last ~7.7 thousand years (ky), without any major population turnover since the early Neolithic.

Southern and northern genetic material in the Japanese and the Koreans

The close genetic affinity between Devil's Gate and modern Japanese and Koreans, who live further south, is also of interest. It has been argued, based on both archaeological (21) and genetic analyses (22–25), that modern Japanese have a dual origin, descending from an admixture event between hunter-gatherers of the Jomon culture (16 to 3 ka) and migrants of the Yayoi culture (3 to 1.7 ka), who brought wet rice agriculture from the Yangtze estuary in southern China through Korea. The few ancient mtDNA samples available from Jomon sites on the northern Hokkaido island show an enrichment of particular haplotypes (N9b and M7a, with D1, D4, and G1 also detected) present in modern Japanese populations, particularly the Ainu and Ryukyuan, as well as southern Siberians (for example, Udegey and Ulchi) (26, 27). The mtDNA haplogroups of our samples from Devil's Gate (D4 and M) are also present in Jomon samples, although they are not the most common ones (N9b and M7a). Recently, nuclear genetic data from two Jomon samples also confirmed the dual origin hypothesis and implied that the Jomon diverged before the diversification of present-day East Asians (28).

We investigated whether it was possible to recover the Northern and Southern genetic components by modeling modern Japanese as a mixture of all possible pairs of sources, including both modern Asian populations and Devil's Gate, using admixture f_3 statistics. The clearest signal was given by a combination of Devil's Gate and modern-day populations from Taiwan, southern China, and Vietnam (Fig. 4), which could represent hunter-gatherer and agriculturalist components, respectively. However, it is important to note that these scores were just barely significant ($-3 < z < -2$) and that some modern pairs also gave



Downloaded from <http://advances.sciencemag.org/> on December 10, 2019

Fig. 1. Regional reference panel, PCA, and ADMIXTURE analysis. (A) Map of Asia showing the location of Devil's Gate (black triangle) and of modern populations forming the regional panel of our analysis. **(B)** Plot of the first two principal components as defined by our regional panel of modern populations from East Asia and central Asia shown on (A), with the two samples from Devil's Gate (black triangles) projected upon them (18). **(C)** ADMIXTURE analysis (19) performed on Devil's Gate and our regional panel, for $K=5$ (lowest cross-validation error) and $K=8$ (appearance of Devil's Gate-specific cluster).

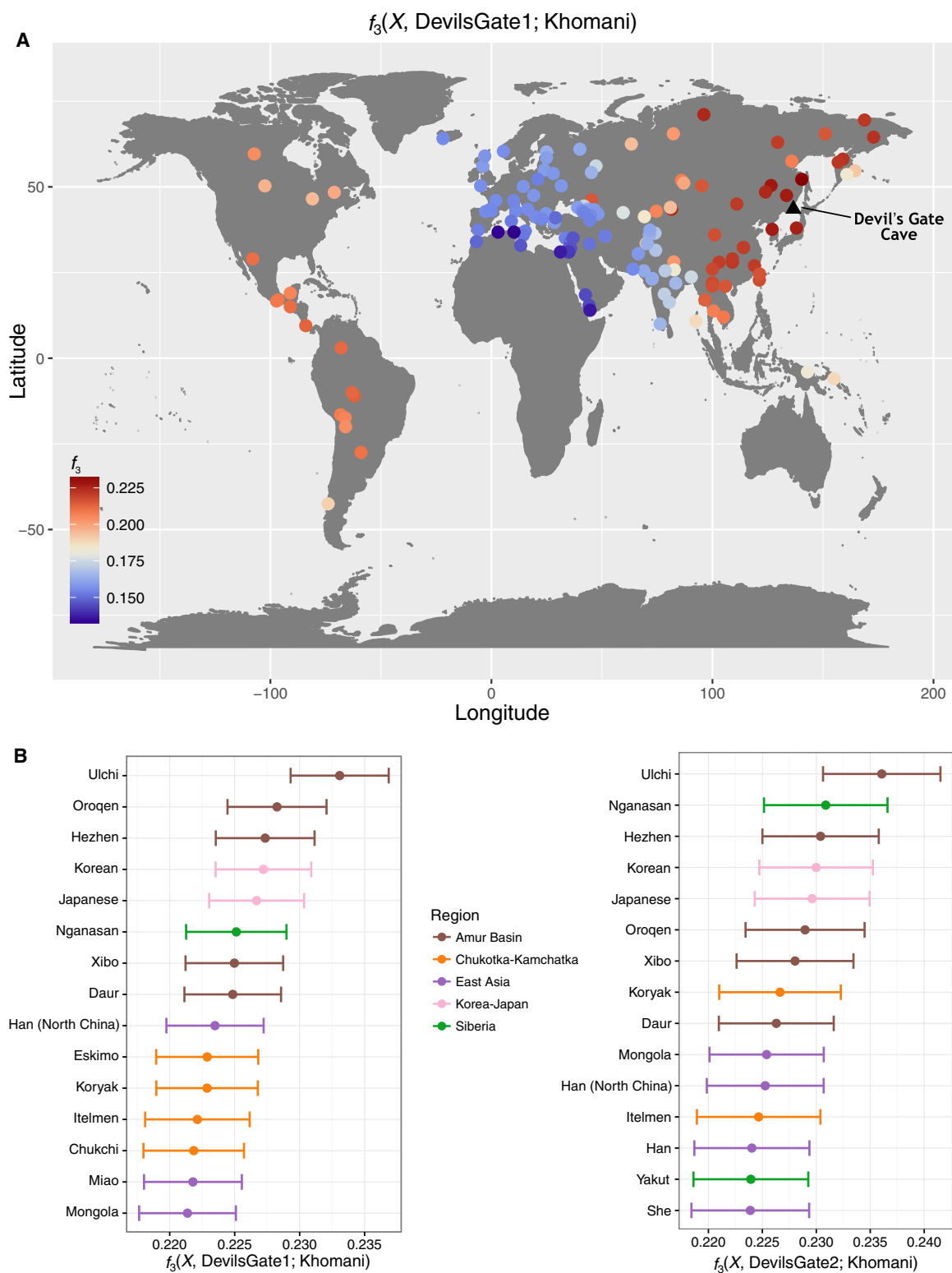


Fig. 2. Outgroup f_3 statistics. Outgroup f_3 measuring shared drift between Devil's Gate (black triangle shows sampling location) and modern populations with respect to an African outgroup (Khomani). (A) Map of the whole world. (B) Fifteen populations with the highest shared drift with Devil's Gate, color-coded by regions as in Fig. 1. Error bars represent 1 SE.

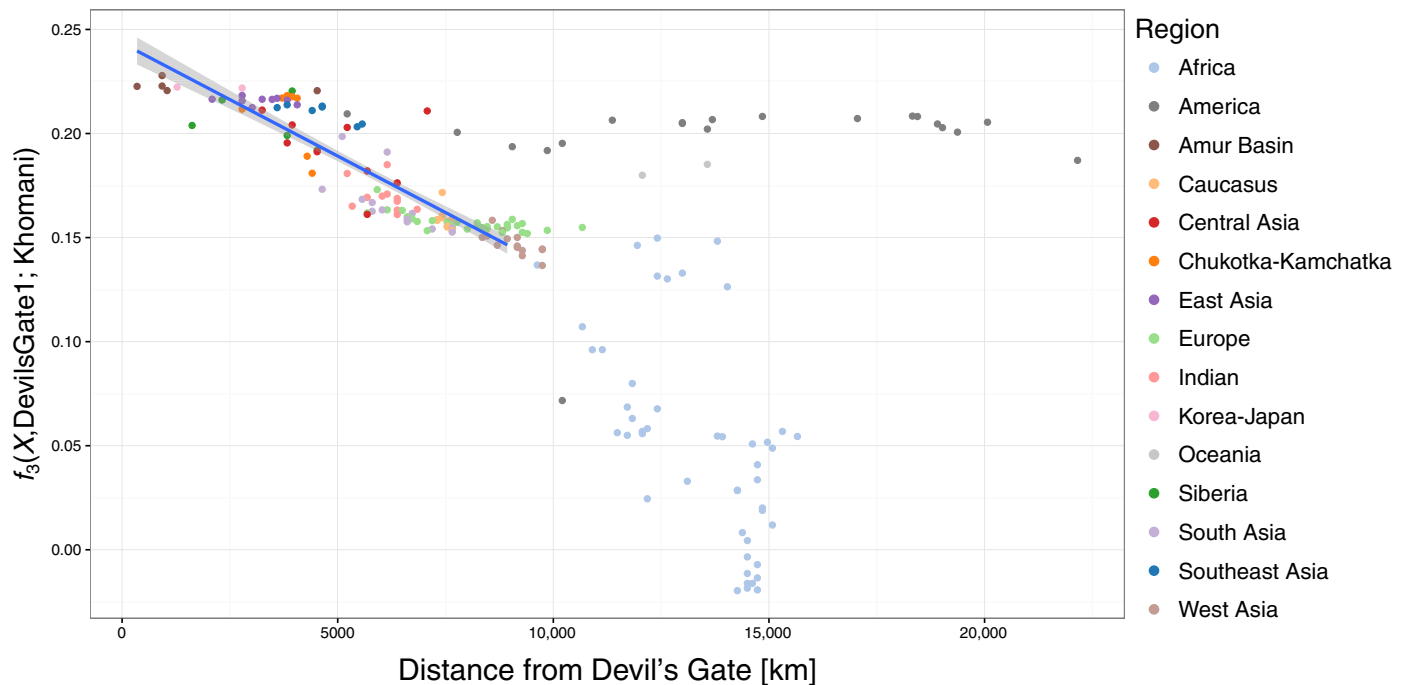


Fig. 3. Spatial pattern of outgroup f_3 statistics. Relationship between outgroup $f_3(X, \text{Devil's Gate}; \text{Khomani})$ and distance on land from Devil's Gate using DevilsGate1 and all single-nucleotide polymorphisms (SNPs). Populations up to 9000 km away from Devil's Gate were considered when computing correlation. The highest distance considered was chosen to acquire the highest Pearson correlation in steps of 500 km. Best linear fit ($r^2 = 0.772$, $F_{1,108} = 368.4$, $P < 0.001$) is shown as blue line, with 95% CI indicated by the shaded area.

negative scores, even if not reaching our significance threshold (z scores as low as -1.9 ; see the Supplementary Materials). The origin of Koreans has received less attention. Also, because of their location on the mainland, Koreans have likely experienced a greater degree of contact with neighboring populations throughout history. However, their genomes show similar characteristics to those of the Japanese on genome-wide SNP data (29) and have also been shown to harbor both northern and southern Asian mtDNA (30) and Y chromosomal haplogroups (30, 31). Unfortunately, our low coverage and small sample size from Devil's Gate prevented a reliable estimate of admixture coefficients or use of linkage disequilibrium-based methods to investigate whether the components originated from secondary contact (admixture) or continuous differentiation and to date any admixture event that did occur.

Phenotypes of interest

The low coverage of our sample does not allow for direct observation of most SNPs linked to phenotypic traits of interest, but imputation based on modern-day populations can provide some information. We focused on the genome with highest coverage, DevilsGate1, using the same imputation approach that has previously been used to estimate genotype probabilities (GPs) for ancient European samples (5, 16, 17). DevilsGate1 likely had brown eyes (rs12913832 on *HERC2*; GP, 0.905) and, where it could be determined, had pigmentation-associated variants that are common in East Asia (see section S11) (32). She appears to have at least one copy of the derived mutation on the *EDAR* gene, encoding the Ectodysplasin A receptor (rs3827760; GP, 0.865), which gives increased odds of straight, thick hair (33), as well as shovel-shaped incisors (34). She almost certainly lacked the most common Eurasian mutation for lactose tolerance (rs4988235, *LCT* gene; GP > 0.999) (35) and was unlikely to have suffered from alcohol flush (rs671, *ALDH2* gene;

GP, 0.847) (36). Thus, at least with regard to those phenotypic traits for which the genetic basis is known, there also seems to have been some degree of phenotypic continuity in this region for the last 7.7 ky.

DISCUSSION

By analyzing genome-wide data from two early Neolithic East Asians from Devil's Gate, in the Russian Far East, we could demonstrate a high level of genetic continuity in the region over at least the last 7700 years. The cold climatic conditions in this area, where modern populations still rely on a number of hunter-gatherer-fisher practices, likely provide an explanation for the apparent continuity and lack of major genetic turnover by exogenous farming populations, as has been documented in the case of southeast and central Europe. Thus, it seems plausible that the local hunter-gatherers progressively added food-producing practices to their original lifestyle. However, it is interesting to note that in Europe, even at very high latitudes, where similar subsistence practices were still important until very recent times, the Neolithic expansion left a significant genetic signature, albeit attenuated in modern populations, compared to the southern part of the continent. Our ancient genomes thus provide evidence for a qualitatively different population history during the Neolithic transition in East Asia compared to western Eurasia, suggesting stronger genetic continuity in the former region. These results encourage further study of the East Asian Neolithic, which would greatly benefit from genetic data from early agriculturalists (ideally, from areas near the origin of wet rice cultivation in southern East Asia), as well as higher-coverage hunter-gatherer samples from different regions to quantify population structure before intensive agriculture.

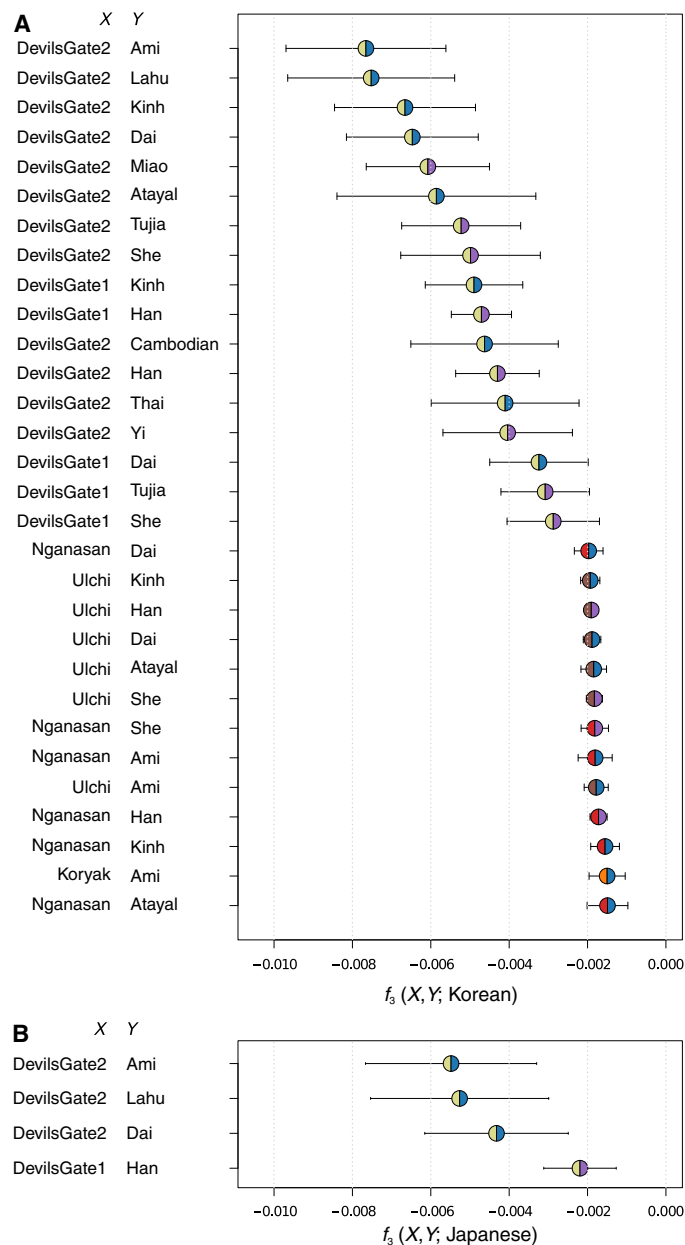


Fig. 4. Admixture f_3 statistics. Admixture f_3 representing modern Koreans and Japanese as a mixture of two populations, X and Y , color-coded by regions as in Fig. 1. **(A)** Thirty pairs with the lowest f_3 score for the Koreans as the target, out of those giving a significant ($z < -2$) value. **(B)** All four pairs giving a significantly ($z < -2$) negative score for the Japanese as the target. Error bars represent 1 SE.

MATERIALS AND METHODS

Experimental design

Sample preparation and sequencing.

Molecular analyses were carried out in dedicated ancient DNA facilities at Trinity College Dublin, Ireland. Samples were prepared, and DNA was extracted using a silica column-based protocol following the methods of Gamba *et al.* (17), which were based on the study of Yang *et al.* (37). DNA extracted from both the first and second lysis buffers (17) was used for library preparation, which was carried out using a modified version of the protocol of Meyer and Kircher (38),

as described by Gamba *et al.* (17). Libraries were sequenced on either an Illumina MiSeq or HiSeq platform (for further details and sequencing statistics, see Supplementary Materials and Methods).

Data processing and alignment.

For single-end sequencing data, adapter sequences were trimmed from the ends of reads using cutadapt (39), allowing an overlap of only 1 base pair (bp) between the adapter and the read. For paired-end data, adapters were trimmed using leeHom (40). leeHom was run using the -ancientdna option, and paired-end reads that overlapped were merged. For paired-end reads that could not be overlapped, only data from read 1 were used in downstream analyses. Reads were aligned using the Burrows-Wheeler alignment tool (BWA) (41), with the seed region disabled, to the GRCh37 build of the human genome, with the mitochondrial sequence replaced by the revised Cambridge reference sequence (National Center for Biotechnology Information accession number NC_012920.1). Reads from different sequencing experiments were merged using Picard MergeSamFiles (<http://picard.sourceforge.net/>), and clonal reads were removed using SAMtools (42). A minimum read length of 30 bp was imposed, and for the higher-coverage (above 0.01X) samples, DevilsGate1 and DevilsGate2, indels were realigned using RealignerTargetCreator and IndelRealigner from the Genome Analysis Toolkit (GATK) (43). SAMtools (42) was used to filter out reads with a mapping quality of less than 30, and reads were rescaled using mapDamage 2.0 (44) to reduce the qualities of likely damaged bases, therefore lessening the effects of ancient DNA damage-associated errors on analysis (44). Average genomic depth of coverage was calculated using the genomecov function of bedtools (45).

Authenticity of results and contamination estimates.

Patterns of molecular damage and the length distribution of reads were assessed using all reads for DevilsGate3, DevilsGate4, and DevilsGate5. Because a portion of the reads from DevilsGate1 and DevilsGate2 was derived from 50-bp single-end sequencing, only reads sequenced with 150-bp paired-end sequencing were considered in the following analyses 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. We replicated all results using our data without applying MapDamage to avoid biases from dropping true mutations that look like damage on our low-coverage data. For further details, see Supplementary Materials and Methods.

We assessed the rate of mitochondrial contamination for our highest-coverage samples, DevilsGate1 and DevilsGate2. This was calculated by evaluating the percentage of nonconsensus bases at haplogroup-defining positions (haplogroup D4 for DevilsGate1 and M for DevilsGate2) using bases with quality ≥ 20 . We also used schmutzi, a tool that uses a Bayesian maximum a posteriori algorithm (12), to estimate the mitochondrial contamination for DevilsGate1. Last, we replicated the analyses that yielded the most robust results (outgroup f_3 scores and PCA) using only reads showing evidence of postmortem damage (PMD score of at least 3) (13). For further details, see Supplementary Materials and Methods.

Statistical analysis

Mitochondrial haplogroup determination and molecular sex determination.

Mitochondrial consensus sequences were generated for DevilsGate1 and DevilsGate2 using Analysis of Next Generation Sequencing Data (ANGSD) (46). Called positions were required to have a depth of coverage ≥ 3 , and only bases with quality ≥ 20 were considered. The resulting

Downloaded from <http://advances.sciencemag.org/> on December 10, 2019

FASTA files were uploaded to HAPLOFIND (47) for haplogroup determination, with coverage calculated using GATK DepthOfCoverage (43). Mutations defining the assigned haplogroup were also manually checked. Molecular sex was assigned using the script described in the study of Skoglund *et al.* (48). For further details, see Supplementary Materials and Methods.

SNP calling and merging with reference panel.

To compare our sample to modern and ancient human genetic variation, we called SNPs using the hg19 reference FASTA file at positions overlapping with the Human Origins (HO) reference panel (591,356 positions) (49) using SAMtools 1.2 (42). Bases were required to have a minimum mapping quality of 30 and base quality of 20; all triallelic SNPs were discarded. Because our low coverage does not provide sufficient information to infer diploid genotypes, a base was chosen with probability proportional to its depth of coverage. This allele was duplicated to form a homozygous diploid genotype, which was used to represent the individual at that SNP position (48). This method of SNP calling (referred to as the proportional method from now on) will artificially increase the appearance of drift on the lineage leading to the ancient individual; however, this drift is not expected to be in any particular direction and, therefore, should not bias inferences about population relationships (3). A total of 35,903 positions in DevilsGate1 and 14,739 positions in DevilsGate2 were covered by at least one high-quality read.

The resulting SNP data for DevilsGate1 and DevilsGate2 were then merged with a reference panel containing modern genomes from the HO panel and selected ancient genomes [this data set was described by Jones *et al.* (5)] as well as an additional 45 Korean genomes from the Personal Genome Project Korea (<http://opengenome.net/>) using PLINK 1.07 (50). Additional sample information is available in extended data table S1, including sample IDs, populations, and groupings used throughout the article. Last, a transversion-only version of all the above data was created by converting all T's to C's and G's to A's. This alternative data set was used to confirm that potential biases originating from ancient DNA damage do not influence our conclusions.

In the later analyses (outgroup and admixture f_3 statistics, PCA, and ADMIXTURE analysis), results using all mutations or only transversions were qualitatively similar, apart from increased noise in the transversion-only data due to the reduced information content. Thus, in the main text, we only report results using all mutations and the default calling method referred to as the proportional method (choosing a read uniformly at random from the reads covering any given position). We present results using the Khomani San as our African outgroup for outgroup f_3 and D statistics, but other populations (the Yoruba, the Mbuti, and the Dinka) gave equivalent results. For further details, see Supplementary Materials and Methods.

Population genetic analysis.

PCA was performed with two different reference panels (Fig. 1 and extended data table S1), both subsets of the worldwide panel of contemporary and ancient individuals from the study of Jones *et al.* (5). The analysis was carried out using EIGENSOFT 6.0.1 smartpca (18), with the lsqproject and normalization 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. For further details and results, see Supplementary Materials and Methods.

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). $K = 2$ to 20 clusters for the global panel and $K = 2$ to 10 clusters for the regional panel were explored 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 and $K = 5$ for the regional (East Asia and central Asia) panel, but the error already started plateauing around $K = 9$ for the global panel, suggesting little improvement. Furthermore, results for the regional panel were largely similar to those from the global panel for East Asian populations. To compare population-level frequencies of the inferred ancestry components, we simultaneously performed bootstrapping on SNPs and individuals for each population, with 100 bootstrap estimates, using inferred ancestral components from our ADMIXTURE runs on all SNPs and MapDamage-treated samples from Devil's Gate. We projected each bootstrap replicate on the ancestral components by numerically maximizing the corresponding likelihood function, following the logic in the study of Allentoft *et al.* (16) and Sikora *et al.* (51). For further details and results, see Supplementary Materials and Methods and extended data figs. S1 to S8.

D statistics (52) and f_3 statistics (49, 53) were used to formally assess the relationships between the samples using the qpDstat (D statistics) and qp3PopTest (f_3 statistics) programs from the ADMIXTOOLS package (49). Significance was assessed by these programs using a block jackknife over 5-centimorgan chunks of the genome, and statistics were considered significant if their z score was of magnitude greater than 2 or, for admixture f_3 scores, if they were smaller than -2 . These correspond approximately to P values of 0.046 and 0.023, respectively. For further details and results, see Supplementary Materials and Methods.

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. Because of the low quality of our samples, we used BEAGLE (54) to impute genotypes using a reference panel containing phased genomes from the 1000 Genomes Project (26 different populations). Following the study of Gamba *et al.* (17), GATK UnifiedGenotyper (43) was used to call genotype likelihoods at SNP sites in phase 3 of the 1000 Genomes Project. Equal likelihoods were set for positions with no spanning sequence data and positions where the observed genotype could be explained by deamination (17). We imputed at least 1 Mb upstream and downstream from the loci of interest using 10 iterations to estimate genotypes at markers for which we had no direct genotype. For further details and results, see Supplementary Materials and Methods and extended data table S9.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/3/2/e1601877/DC1>

Supplementary Materials and Methods

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.

fig. S3. Sequence length distribution for samples from Devil's Gate.

fig. S4. Outgroup f_3 statistics on PMDtools-filtered data.

fig. S5. PCA on all SNPs using the worldwide panel.

fig. S6. PCA on transversion SNPs using the worldwide panel.

fig. S7. PCA on all SNPs using the regional panel.

fig. S8. PCA on transversion SNPs using the regional panel.

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.

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.

fig. S11. Outgroup f_3 scores of the form $f_3(X, MA1; Khomani)$, with modern populations and selected ancient samples (DevilsGate1, DevilsGate2, Ust'-Ishim, Kotias, Loschbour, and Stuttgart), using all SNPs, with $f_3 > 0.15$ displayed.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

extended data table S1. Sample information.

extended data table S2. ADMIXTURE proportions.

extended data table S3. Outgroup f_3 statistics for Devil's Gate.

extended data table S4. Outgroup f_3 and space.

extended data table S5. Outgroup f_3 for MA1 and Ust'-Ishim.

extended data table S6. D scores for MA1 and Ust'-Ishim.

extended data table S7. D scores for the Ulchi.

extended data table S8. Admixture f_3 for the Koreans and the Japanese.

extended data table S9. Phenotypes of interest.

References (55–84)

REFERENCES AND NOTES

1. I. Lazaridis, D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland, S. Mallick, D. Fernandes, M. Novak, B. Gamarra, K. Sirak, S. Connell, K. Stewardson, E. Harnay, Q. Fu, G. Gonzalez-Fortes, E. R. Jones, S. Alpaslan Roodenberg, G. Lengyel, F. Bocquentin, B. Gasparian, J. M. Monge, M. Gregg, V. Eshed, A.-S. Mizrahi, C. Meiklejohn, F. Gerritsen, L. Bejenaru, M. Blüher, A. Campbell,

- G. Cavalleri, D. Comas, P. Froguel, E. Gilbert, S. M. Kerr, P. Kovacs, J. Krause, D. McGettigan, M. Merrigan, D. Andrew Merriwether, S. O'Reilly, M. B. Richards, O. Semino, M. Shamoony-Pour, G. Stefanescu, M. Stumvoll, A. Tönjes, A. Torroni, J. F. Wilson, L. Yengo, N. A. Hovhannisyun, N. Patterson, R. Pinhasi, D. Reich, Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
2. M. Gallego-Llorente, S. Connell, E. R. Jones, D. C. Merrett, Y. Jeon, A. Eriksson, V. Siska, C. Gamba, C. Meiklejohn, R. Beyer, S. Jeon, Y. S. Cho, M. Hofreiter, J. Bhak, A. Manica, R. Pinhasi, The genetics of an early Neolithic pastoralist from the Zagros, Iran. *Sci. Rep.* **6**, 31326 (2016).
3. I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfelt, H. Li, C. de Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J.-M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovsky, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. J. Busby, F. Cali, M. Churmosov, D. E. C. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J.-M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kučinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Näkkäljärvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekgn, D. Toncheva, S. Turdikulova, I. Uktvertye, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villem, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Pääbo, J. Kelso, D. Reich, J. Krause, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
4. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harnay, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. Garrido Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. Werner Alt, D. Reich, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
5. E. R. Jones, G. Gonzalez-Fortes, S. Connell, V. Siska, A. Eriksson, R. Martiniano, R. L. McLaughlin, M. Gallego Llorente, L. M. Cassidy, C. Gamba, T. Meshveliani, O. Bar-Yosef, W. Müller, A. Belfer-Cohen, Z. Matskevich, N. Jakeli, T. F. G. Higham, C. Currat, D. Lordkipanidze, M. Hofreiter, A. Manica, R. Pinhasi, D. G. Bradley, Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
6. M. Raghavan, M. DeGiorgio, A. Albrechtsen, I. Moltke, P. Skoglund, T. S. Korneliussen, B. Gronnow, M. Appelt, H. Christian Gulløv, T. Max Friesen, W. Fitzhugh, H. Malmström, S. Rasmussen, J. Olsen, L. Melchior, B. T. Fuller, S. M. Fahrni, T. Stafford Jr., V. Grimes, M. A. P. Renouf, J. Cybulski, N. Lynnerup, M. M. Lahr, K. Britton, R. Knecht, J. Arneborg, M. Metspalu, O. E. Cornejo, A.-S. Malaspinas, Y. Wang, M. Rasmussen, V. Raghavan, T. V. O. Hansen, E. Khusnutdinova, T. Pierre, K. Dneprovsky, C. Andreasen, H. Lange, M. G. Hayes, J. Coltrain, V. A. Spitsyn, A. Götherström, L. Orlando, T. Kivisild, R. Villem, M. H. Crawford, F. C. Nielsen, J. Dissing, J. Heinemeier, M. Meldgaard, C. Bustamante, D. H. O'Rourke, M. Jakobsson, M. T. P. Gilbert, R. Nielsen, E. Willerslev, The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014).
7. M. Raghavan, M. Steinrücken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, A. Albrechtsen, C. Valdiosera, M. C. Ávila-Arcos, A.-S. Malaspinas, A. Eriksson, I. Moltke, M. Metspalu, J. R. Homburger, J. Wall, O. E. Cornejo, J. Víctor Moreno-Mayar, T. S. Korneliussen, T. Pierre, M. Rasmussen, P. F. Campos, P. de Barros Damgaard, M. E. Allentoft, J. Lindo, E. Metspalu, R. Rodríguez-Varela, J. Mansilla, C. Henriksson, A. Seguin-Orlando, H. Malmström, T. Stafford Jr., S. S. Shringarpure, A. Moreno-Estrada, M. Karmin, K. Tambets, A. Bergström, Y. Xue, V. Warmuth, A. D. Friend, J. Singarayer, P. Valdes, F. Balloux, I. LeBoreiro, J. L. Vera, H. Rangel-Villalobos, D. Pettener, D. Luiselli, L. G. Davis, E. Heyer, C. P. E. Zollikofer, M. S. Ponce de León, C. I. Smith, V. Grimes, K.-A. Pike, M. Deal, B. T. Fuller, B. Arriaza, V. Standen, M. F. Luz, F. Ricaut, N. Guidon, L. Osipova, M. I. Voevoda, O. L. Posukh, O. Balanovsky, M. Lavryashina, Y. Bogunov, E. Khusnutdinova, M. Gubina, E. Balanovska, S. Fedorova, S. Litvinov, B. Malyarchuk, M. Derenko, M. J. Mosher, D. Archer, J. Cybulski, B. Petzelt, J. Mitchell, R. Worl, P. J. Norman, P. Parham, B. M. Kemp, T. Kivisild, C. Tyler-Smith, M. S. Sandhu, M. Crawford, R. Villem, D. G. Smith, M. R. Waters, T. Goebel, J. R. Johnson, R. S. Malhi, M. Jakobsson, D. J. Meltzer, A. Manica, R. Durbin, C. D. Bustamante, Y. S. Song, R. Nielsen, E. Willerslev, Genomic evidence for the Pleistocene and recent population history of native Americans. *Science* **349**, aab3884 (2015).
8. Y. Kuzmin, The Paleolithic-to-Neolithic transition and the origin of pottery production in the Russian Far East: A geoarchaeological approach. *Archaeol. Ethnol. Anthropol. Eurasia* **4**, 16–26 (2003).
9. Z. Chi, H. Hsiao-chun, *The Encyclopedia of Global Human Migration* (Blackwell Publishing Ltd., 2013).

10. Y. V. Kuzmin, C. T. Keally, A. J. T. Jull, G. S. Burr, N. A. Klyuev, The earliest surviving textiles in East Asia from Chertovy Vorota Cave, Primorye province, Russian Far East. *Antiquity* **86**, 325–337 (2012).
11. M. Tanaka, V. M. Cabrera, A. M. González, J. M. Larruga, T. Takeyasu, N. Fuku, L.-J. Guo, R. Hirose, Y. Fujita, M. Kurata, K.-i. Shinoda, K. Umetsu, Y. Yamada, Y. Oshida, Y. Sato, N. Hattori, Y. Mizuno, Y. Arai, N. Hirose, S. Ohta, O. Ogawa, Y. Tanaka, R. Kawamori, M. Shamoto-Nagai, W. Maruyama, H. Shimokata, R. Suzuki, H. Shimodaira, Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res.* **14**, 1832–1850 (2004).
12. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
13. P. Skoglund, B. H. Northoff, M. V. Shunkov, A. P. Derevianko, S. Pääbo, J. Krause, M. Jakobsson, Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2229–2234 (2014).
14. Q. Fu, H. Li, P. Moorjani, F. Jay, S. M. Slepchenko, A. A. Bondarev, P. L. F. Johnson, A. Aximu-Petri, K. Prüfer, C. de Filippo, M. Meyer, N. Zwyns, D. C. Salazar-García, Y. V. Kuzmin, S. G. Keates, P. A. Kosintsev, D. I. Razhev, M. P. Richards, N. V. Peristov, M. Lachmann, K. Douka, T. F. G. Higham, M. Slatkin, J.-J. Hublin, D. Reich, J. Kelso, T. Benice Viola, S. Pääbo, Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
15. M. Raghavan, P. Skoglund, K. E. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, S. Rasmussen, T. W. Stafford Jr., L. Orlando, E. Metspalu, M. Karmin, K. Tambets, S. Rootsi, R. Mägi, P. F. Campos, E. Balanovska, O. Balanovsky, E. Khusnutdinova, S. Litvinov, L. P. Osipova, S. A. Fedorova, M. I. Voevoda, M. DeGiorgio, T. Sicheritz-Ponten, S. Brunak, S. Demeshchenko, T. Kivisild, R. Villems, R. Nielsen, M. Jakobsson, E. Willerslev, Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
16. M. E. Allentoft, M. Sikora, K.-G. Sjögren, S. Rasmussen, M. Rasmussen, J. Stenderup, P. B. Damgaard, H. Schroeder, T. Ahlström, L. Vinner, A.-S. Malaspina, A. Margaryan, T. Higham, D. Chivall, N. Lynnerup, L. Harvig, J. Baron, P. Della Casa, P. Dąbrowski, P. R. Duffy, A. V. Ebel, A. Epimakhov, K. Frei, M. Furmanek, T. Gralak, A. Gromov, S. Gronkiewicz, G. Grupe, T. Hajdu, R. Jarysz, V. Khartanovich, A. Khokhlov, V. Kiss, J. Kolář, A. Kriska, I. Lasak, C. Longhi, G. McGlynn, A. Merkevicius, I. Merkyte, M. Metspalu, R. Mkrtychyan, V. Moiseyev, L. Paja, G. Pálfi, D. Pokutta, L. Pospieszny, T. Douglas Price, L. Saag, M. Sablin, N. Shishlina, V. Smrčka, V. I. Soenov, V. Szeverényi, G. Tóth, S. V. Trifanov, L. Varul, M. Vicze, L. Yepiskoposyan, V. Zhitenev, L. Orlando, T. Sicheritz-Pontén, S. Brunak, R. Nielsen, K. Kristiansen, E. Willerslev, Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
17. C. Gamba, E. R. Jones, M. D. Teasdale, R. L. McLaughlin, G. Gonzalez-Fortes, V. Mattiangeli, L. Domboróczki, I. Kóvári, I. Pap, A. Anders, A. Whittle, J. Dani, P. Raczky, T. F. G. Higham, M. Hofreiter, D. G. Bradley, R. Pinhasi, Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
18. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006).
19. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **9**, 1655–1664 (2009).
20. F. Sánchez-Quinto, H. Schroeder, O. Ramirez, M. C. Ávila-Arcos, M. Pybus, I. Olalde, A. M. V. Velazquez, M. E. P. Marcos, J. M. V. Encinas, J. Bertranpetit, L. Orlando, M. T. P. Gilbert, C. Laluzza-Fox, Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr. Biol.* **22**, 1494–1499 (2012).
21. K. Imamura, *Prehistoric Japan: New Perspectives on Insular East Asia* (Univ. Hawaii Press, 1996).
22. T. A. Jinam, H. Kanzawa-Kiriyama, N. Saitou, Human genetic diversity in the Japanese Archipelago: Dual structure and beyond. *Genes Genet. Syst.* **90**, 147–152 (2015).
23. R. Rasteiro, L. Chikhi, Revisiting the peopling of Japan: An admixture perspective. *J. Hum. Genet.* **54**, 349–354 (2009).
24. Y. He, W. R. Wang, S. Xu, L. Jin; Pan-Asia SNP Consortium, Paleolithic contingent in modern Japanese: Estimation and inference using genome-wide data. *Sci. Rep.* **2**, 355 (2012).
25. T. A. Jinam, H. Kanzawa-Kiriyama, I. Inoue, K. Tokunaga, K. Omoto, N. Saitou, Unique characteristics of the Ainu population in Northern Japan. *J. Hum. Genet.* **60**, 565–571 (2015).
26. N. Adachi, K.-i. Shinoda, K. Umetsu, H. Matsumura, Mitochondrial DNA analysis of Jomon skeletons from the Funadomari site, Hokkaido, and its implication for the origins of Native American. *Am. J. Phys. Anthropol.* **138**, 255–265 (2009).
27. N. Adachi, K.-i. Shinoda, K. Umetsu, T. Kitano, H. Matsumura, R. Fujiyama, J. Sawada, M. Tanaka, Mitochondrial DNA analysis of Hokkaido Jomon skeletons: Remnants of archaic maternal lineages at the southwestern edge of former Beringia. *Am. J. Phys. Anthropol.* **146**, 346–360 (2011).
28. H. Kanzawa-Kiriyama, K. Kryukov, T. A. Jinam, K. Hosomichi, A. Saso, G. Suwa, S. Ueda, M. Yoneda, A. Tajima, K.-i. Shinoda, I. Inoue, N. Saitou, A partial nuclear genome of the Jomons who lived 3000 years ago in Fukushima, Japan. *J. Hum. Genet.* 10.1038/jhg.2016.110 (2016).
29. HUGO Pan-Asian SNP Consortium, Mapping human genetic diversity in Asia. *Science* **326**, 1541–1545 (2009).
30. H.-J. Jin, C. Tyler-Smith, W. Kim, The peopling of Korea revealed by analyses of mitochondrial DNA and Y-chromosomal markers. *PLOS ONE* **4**, e4210 (2009).
31. H.-J. Jin, K.-D. Kwak, M. F. Hammer, Y. Nakahori, T. Shinka, J.-W. Lee, F. Jin, X. Jia, C. Tyler-Smith, W. Kim, Y-chromosomal DNA haplogroups and their implications for the dual origins of the Koreans. *Hum. Genet.* **114**, 27–35 (2003).
32. R. A. Sturm, D. L. Duffy, Human pigmentation genes under environmental selection. *Genome Biol.* **13**, 248 (2012).
33. A. Fujimoto, R. Kimura, J. Ohashi, K. Omi, R. Yuliwulandari, L. Batubara, M. S. Mustofa, U. Samakkarn, W. Settheetham-Ishida, T. Ishida, Y. Morishita, T. Furusawa, M. Nakazawa, R. Ohtsuka, K. Tokunaga, A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum. Mol. Genet.* **17**, 835–843 (2008).
34. J.-H. Park, T. Yamaguchi, C. Watanabe, A. Kawaguchi, K. Haneji, M. Takeda, Y.-I. Kim, Y. Tomoyasu, M. Watanabe, H. Oota, T. Hanihara, H. Ishida, K. Maki, S.-B. Park, R. Kimura, Effects of an Asian-specific nonsynonymous EDAR variant on multiple dental traits. *J. Hum. Genet.* **57**, 508–514 (2012).
35. Y. Itan, A. Powell, M. A. Beaumont, J. Burger, M. G. Thomas, The origins of lactase persistence in Europe. *PLOS Comput. Biol.* **5**, e1000491 (2009).
36. D. W. Crabb, H. J. Edenberg, W. F. Bosron, T. K. Li, Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive ALDH2² allele is dominant. *J. Clin. Invest.* **83**, 314–316 (1989).
37. D. Y. Yang, B. Eng, J. S. Wayne, J. C. Dudar, S. R. Saunders, Improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* **105**, 539–543 (1998).
38. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
39. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**, 10–12 (2011).
40. G. Renaud, U. Stenzel, J. Kelso, leeHm: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).
41. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
42. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
43. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
44. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
45. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
46. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
47. D. Vianello, F. Sevinci, G. Castellani, L. Lomartire, M. Capri, C. Franceschi, HAPLOFIND: A new method for high-throughput mtDNA haplogroup assignment. *Hum. Mutat.* **34**, 1189–1194 (2013).
48. P. Skoglund, H. Malmström, A. Omrak, M. Raghavan, C. Valdiosera, T. Günther, P. Hall, K. Tambets, J. Parik, K.-G. Sjögren, J. Apel, E. Willerslev, J. Storå, A. Götherström, M. Jakobsson, Genomic diversity and admixture differs for stone-age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014).
49. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webber, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
50. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
51. M. Sikora, M. L. Carpenter, A. Moreno-Estrada, B. M. Henn, P. A. Underhill, F. Sánchez-Quinto, I. Zaza, M. Pitzalis, C. Sidore, F. Busonero, A. Maschio, A. Angius, C. Jones, J. Mendoza-Revilla, G. Nekhrizov, D. Dimitrova, N. Theodosios, T. T. Harkins, A. Keller, F. Mäxner, A. Zink, G. Abecasis, S. Sanna, F. Cucca, C. D. Bustamante, Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean iceman and the genetic structure of Europe. *PLOS Genet.* **10**, e1004353 (2014).
52. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, N. F. Hansen, E. Y. Durand, A.-S. Malaspina, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höflner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Auffrith, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušić, V. B. Doronichev, L. V. Golovanova, C. Laluzza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
53. D. Reich, K. Thangaraj, N. Patterson, A. L. Price, L. Singh, Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).

54. B. L. Browning, S. R. Browning, Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).
55. Z. V. Andreeva, V. Tatarnikov, Peshera "Chertovoy Vorota v Primor'e". *Arheol. Otkrytiya 1973 Goda Cave Devils Gate Primorye 1973 Excav. Seas. Mosc.*, 180–181 (1974).
56. I. S. Zhushchikhovskaya, *Archaeology of the Russian Far East: Essays in Stone Age Prehistory*, S. M. Nelson, A. P. Derevianko, Y. V. Kuzmin, R. L. Bland, Eds. (Archaeopress, 2006), pp. 101–122.
57. A. Kuznecov, The ancient settlement in a cave Devil's Gate and some problems of the Neolithic of Primorye. *Russ. Arheol.* **2**, 17–19 (2002).
58. T. S. Balueva, Cranial samples from a Neolithic layer from Devil's Gate Cave, Primorye. *Vopr. Antropol.* **58**, 184–187 (1978).
59. M. G. Llorente, E. R. Jones, A. Eriksson, V. Siska, K. W. Arthur, J. W. Arthur, M. C. Curtis, J. T. Stock, M. Coltroni, P. Pieruccini, S. Stretton, F. Brock, T. Higham, Y. Park, M. Hofreiter, D. G. Bradley, J. Bhak, R. Pinhasi, A. Manica, Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820–822 (2015).
60. B. Shapiro, M. Hofreiter, A paleogenomic perspective on evolution and gene function: New insights from ancient DNA. *Science* **343**, 1236573 (2014).
61. A. Manica, W. Amos, F. Balloux, T. Hanihara, The effect of ancient population bottlenecks on human phenotypic variation. *Nature* **448**, 346–348 (2007).
62. R. L. Lamason, M.-A. P. K. Mohideen, J. R. Mest, A. C. Wong, H. L. Norton, M. C. Aros, M. J. Jurynec, X. Mao, V. R. Humphreville, J. E. Humbert, S. Sinha, J. L. Moore, P. Jagadeeswaran, W. Zhao, G. Ning, I. Makalowska, P. M. McKeigue, O. O'Donnell, R. Kittles, E. J. Parra, N. J. Mangini, D. J. Grunwald, M. D. Shriver, V. A. Canfield, K. C. Cheng, SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786 (2005).
63. R. A. Sturm, Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet.* **18**, R9–R17 (2009).
64. V. A. Canfield, A. Berg, S. Peckins, S. M. Wentzel, K. Chung Ang, S. Oppenheimer, K. C. Cheng, Molecular phylogeography of a human autosomal skin color locus under natural selection. *G3* **3**, 2059–2067 (2013).
65. C. Basu Mallick, F. Mircea Iliescu, M. Möls, S. Hill, R. Tamang, G. Chaubey, R. Goto, S. Y. W. Ho, I. Gallego Romero, F. Crivellaro, G. Hudjashov, N. Rai, M. Metspalu, C. G. N. Mascie-Taylor, R. Pitchappan, L. Singh, M. Mirazon-Lahr, K. Thangaraj, R. Villemis, T. Kivisild, The light skin allele of SLC24A5 in south Asians and Europeans shares identity by descent. *PLoS Genet.* **9**, e1003912 (2013).
66. S. Wilde, A. Timpson, K. Kirsanow, E. Kaiser, M. Kayser, M. Unterländer, N. Hoffelder, I. D. Potekhina, W. Schier, M. G. Thomas, J. Burger, Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4832–4837 (2014).
67. H. L. Norton, R. A. Kittles, E. Parra, P. McKeigue, X. Mao, K. Cheng, V. A. Canfield, D. G. Bradley, B. McEvoy, M. D. Shriver, Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24**, 710–722 (2007).
68. M. P. Donnelly, P. Paschou, E. Grigorenko, D. Gurwitz, C. Barta, R.-B. Lu, O. V. Zhukova, J.-J. Kim, M. Siniscalco, M. New, H. Li, S. L. B. Kajuna, V. G. Manolopoulos, W. C. Speed, A. J. Pakstis, J. R. Kidd, K. K. Kidd, A global view of the OCA2-HERC2 region and pigmentation. *Hum. Genet.* **131**, 683–696 (2011).
69. J. L. Hider, R. M. Gittelman, T. Shah, M. Edwards, A. Rosenbloom, J. M. Akey, E. J. Parra, Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. *BMC Evol. Biol.* **13**, 150 (2013).
70. I. Yuasa, K. Umetsu, S. Harihara, A. Kido, A. Miyoshi, N. Saitou, B. Dashnyam, F. Jin, G. Lucotte, P. K. Chattopadhyay, L. Henke, J. Henke, Distribution of two Asian-related coding SNPs in the MC1R and OCA2 genes. *Biochem. Genet.* **45**, 535–542 (2007).
71. I. Yuasa, K. Umetsu, S. Harihara, A. Miyoshi, N. Saitou, K. Sook Park, B. Dashnyam, F. Jin, G. Lucotte, P. K. Chattopadhyay, L. Henke, J. Henke, OCA2*481Thr, a hypofunctional allele in pigmentation, is characteristic of northeastern Asian populations. *J. Hum. Genet.* **52**, 690–693 (2007).
72. I. Yuasa, S. Harihara, F. Jin, H. Nishimukai, J. Fujihara, Y. Fukumori, H. Takeshita, K. Umetsu, N. Saitou, Distribution of OCA2*481Thr and OCA2*615Arg, associated with hypopigmentation, in several additional populations. *Leg. Med.* **13**, 215–217 (2011).
73. E. M. Ramos, D. Hoffman, H. A. Junkins, D. Maglott, L. Phan, S. T. Sherry, M. Feolo, L. A. Hindorf, Phenotype–Genotype Integrator (PheGenI): Synthesizing genome-wide association study (GWAS) data with existing genomic resources. *Eur. J. Hum. Genet.* **22**, 144–147 (2014).
74. A. Fujimoto, J. Ohashi, N. Nishida, T. Miyagawa, Y. Morishita, T. Tsunoda, R. Kimura, K. Tokunaga, A replication study confirmed the EDAR gene to be a major contributor to population differentiation regarding head hair thickness in Asia. *Hum. Genet.* **124**, 179–185 (2008).
75. Y. G. Kamberov, S. Wang, J. Tan, P. Gerbault, A. Wark, L. Tan, Y. Yang, S. Li, K. Tang, H. Chen, A. Powell, Y. Itan, D. Fuller, J. Lohmueller, J. Mao, A. Schachar, M. Paymer, E. Hostetter, E. Byrne, M. Burnett, A. P. McMahon, M. G. Thomas, D. E. Lieberman, L. Jin, C. J. Tabin, B. A. Morgan, P. C. Sabeti, Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* **152**, 691–702 (2013).
76. H. W. Goedde, D. P. Agarwal, G. Fritze, D. Meier-Tackmann, S. Singh, G. Beckmann, K. Bhatia, L. Z. Chen, B. Fang, R. Lisker, Y. K. Paik, F. Rothhammer, N. Saha, B. Segal, L. M. Srivastava, A. Czeizel, Distribution of ADH2 and ALDH2 genotypes in different populations. *Hum. Genet.* **88**, 344–346 (1992).
77. H. Li, S. Borinskaya, K. Yoshimura, N. Kal'ina, A. Marusin, V. A. Stepanov, Z. Qin, S. Khalig, M.-Y. Lee, Y. Yang, A. Mohyuddin, D. Gurwitz, S. Q. Mehdi, E. Rogav, L. Jin, N. K. Yankovsky, J. R. Kidd, K. K. Kidd, Refined geographic distribution of the oriental ALDH2*504Lys (nee 487Lys) variant. *Ann. Hum. Genet.* **73**, 335–345 (2009).
78. K. Matsuo, N. Hamajima, M. Shinoda, S. Hatoooka, M. Inoue, T. Takezaki, K. Tajima, Gene–environment interaction between an aldehyde dehydrogenase-2 (ALDH2) polymorphism and alcohol consumption for the risk of esophageal cancer. *Carcinogenesis* **22**, 913–916 (2001).
79. D. Cusi, C. Barlassina, T. Azzani, G. Casari, L. Citterio, M. Devoto, N. Glorioso, C. Lanzani, P. Manunta, M. Righetti, R. Rivera, P. Stella, C. Troffa, L. Zagato, G. Bianchi, Polymorphisms of α -adducin and salt sensitivity in patients with essential hypertension. *Lancet* **349**, 1353–1357 (1997).
80. E. van den Wildenberg, R. W. Wiers, J. Dessers, R. G. J. H. Janssen, E. H. Lambrichs, H. J. M. Smeets, G. J. P. van Breukelen, A functional polymorphism of the μ -opioid receptor gene (OPRM1) influences cue-induced craving for alcohol in male heavy drinkers. *Alcohol. Clin. Exp. Res.* **31**, 1–10 (2007).
81. B. S. Haerian, M. S. Haerian, OPRM1 rs1799971 polymorphism and opioid dependence: Evidence from a meta-analysis. *Pharmacogenomics* **14**, 813–824 (2013).
82. S. Rodriguez, C. D. Steer, A. Farrow, J. Golding, I. N. M. Day, Dependence of deodorant usage on ABC11 genotype: Scope for personalized genetics in personal hygiene. *J. Invest. Dermatol.* **133**, 1760–1767 (2013).
83. K.-i. Yoshiura, A. Kinoshita, T. Ishida, A. Ninokata, T. Ishikawa, T. Kaname, M. Bannai, K. Tokunaga, S. Sonoda, R. Komaki, M. Ihara, V. A. Saenko, G. K. Alipov, I. Sekine, K. Komatsu, H. Takahashi, M. Nakashima, N. Sosonkina, C. K. Mapendano, M. Ghadami, M. Nomura, D.-S. Liang, N. Miwa, D.-K. Kim, A. Garidkhuu, N. Natsume, T. Ohta, H. Tomita, A. Kaneko, M. Kikuchi, G. Russomando, K. Hirayama, M. Ishibashi, A. Takahashi, N. Saitou, J. C. Murray, S. Saito, Y. Nakamura, N. Niikawa, A SNP in the ABC11 gene is the determinant of human earwax type. *Nat. Genet.* **38**, 324–330 (2006).
84. Y. S. Cho, C.-H. Chen, C. Hu, J. Long, R. T. H. Ong, X. Sim, F. Takeuchi, Y. Wu, M. Jin Go, T. Yamauchi, Y.-C. Chang, S. H. Kwak, R. C. W. Ma, K. Yamamoto, L. S. Adair, T. Aung, Q. Cai, L.-C. Chang, Y.-T. Chen, Y. Gao, F. B. Hu, H.-L. Kim, S. Kim, Y. J. Kim, J. J.-M. Lee, N. R. Lee, Y. Li, J. J. Liu, W. Lu, J. Nakamura, E. Nakashima, D. P.-K. Ng, W. T. Tay, F.-J. Tsai, T. Y. Wong, M. Yokota, W. Zheng, R. Zhang, C. Wang, W. Y. So, K. Ohnaka, H. Ikegami, K. Hara, Y. M. Cho, N. H. Cho, T.-J. Chang, Y. Bao, Å. K. Hedman, A. P. Morris, M. I. McCarthy; DIAGRAM Consortium; MuTHER Consortium, R. Takayanagi, K. S. Park, W. Jia, L.-M. Chuang, J. C. N. Chan, S. Maeda, T. Kadowaki, J.-Y. Lee, J.-Y. Wu, Y. Y. Teo, E. S. Tai, X. O. Shu, K. L. Mohlke, N. Kato, B.-G. Han, M. Seielstad, Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in East Asians. *Nat. Genet.* **44**, 67–72 (2012).

Acknowledgments

Funding: V.S. was supported by the Gates Cambridge Trust. R.P. was funded by the European Research Council (ERC) starting grant ADNABIOARC (263441) and the Irish Research Council Advanced Research Project Grant from January 2014 to December 2016. M.H. was supported by ERC Consolidator Grant 310763 "GeneFlow." This work was supported by the Research Fund (1.140113.01) of Ulsan National Institute of Science and Technology to J.B. This work was also supported by the Research Fund (14-BR-SS-03) of Civil-Military Technology Cooperation Program to J.B. and Y.S.C. M.G.-L. was supported by a Biotechnology and Biological Sciences Research Council Doctoral Training Partnerships studentship. A.M. and A.E. were supported by the ERC Consolidator Grant 647787 "LocalAdaptation." D.G.B. was funded by ERC Investigator grant 295729-CodeX. **Author contributions:** E.V., T.B., and R.P. acquired the samples and provided the archaeological context; E.R.J., H.-M.K., Y.S.C., H.K., and K.L. performed experiments; V.S., E.R.J., S.J., Y.B., Y.S.C., M.G.-L., J.B., and A.M. analyzed genetic data; V.S., M.H., D.G.B., A.E., R.P., J.B., and A.M. wrote the manuscript with input from all coauthors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. Raw reads in FASTA format and aligned reads as bamfiles for all five ancient samples from Devil's Gate are available from the European Nucleotide Archive accession code PRJEB14817.

Submitted 10 August 2016

Accepted 21 December 2016

Published 1 February 2017

10.1126/sciadv.1601877

Citation: V. Siska, E. R. Jones, S. Jeon, Y. Bhak, H.-M. Kim, Y. S. Cho, H. Kim, K. Lee, E. Veselovskaya, T. Balueva, M. Gallego-Llorente, M. Hofreiter, D. G. Bradley, A. Eriksson, R. Pinhasi, J. Bhak, A. Manica, Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Sci. Adv.* **3**, e1601877 (2017).

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 and Andrea Manica

Sci Adv 3 (2), e1601877.
DOI: 10.1126/sciadv.1601877

ARTICLE TOOLS

<http://advances.sciencemag.org/content/3/2/e1601877>

SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2017/01/30/3.2.e1601877.DC1>

REFERENCES

This article cites 79 articles, 15 of which you can access for free
<http://advances.sciencemag.org/content/3/2/e1601877#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2017, The Authors