Ancient human genomes suggest three ancestral populations for present-day Europeans

Iosif Lazaridis, Nick Patterson, Alissa Mittnik, et al.

bioRxiv posted online December 23, 2013
Access the most recent version at doi: 10.1101/001552
Ancient human genomes suggest three ancestral populations for present-day Europeans


† Co-senior authors.

To whom correspondence should be addressed: David Reich (reich@genetics.med.harvard.edu) or Johannes Krause (johannes.krause@uni-tuebingen.de)
NextBio, part of Illumina, Santa Clara, CA, USA 95050
College of Medicine, University of Illinois at Chicago, Chicago, IL, USA 60607
Institute of Biochemistry and Genetics, Ufa Research Centre, Russian Academy of Sciences, Ufa, Russia 450054
Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia 450074
Division of Biological Anthropology, University of Cambridge, Cambridge, United Kingdom CB2 1QH
Department of Human and Medical Genetics, Vilnius University, Vilnius, Lithuania LT-08661
Estonian Biocentre, Evolutionary Biology group, Tartu, Estonia 51010
Translational Medicine and Neurogenetics, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France 67404
Robert W. Mahley, Gladstone Institutes, San Francisco, CA 94158
Department of Evolutionary Biology, University of Tartu, Tartu, Estonia 51010
Research Department, 23andMe, Inc. Mountain View, CA, USA 94043
Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania
Research Institute of Health, North-Eastern Federal University, Yakutsk, Russia 677000
Dipartimento di Fisica e Chimica, Università di Palermo, Palermo, Italy 90128
Centre for Population Health Sciences, The University of Edinburgh Medical School, Edinburgh, Scotland, United Kingdom EH8 9AG
Deceased: formerly of the Institute of Immunology, Academy of Science, Tashkent, Uzbekistan 70000
Laboratory of Ethnogenomics, Institute of Molecular Biology, National Academy of Sciences of Armenia, Yerevan, Armenia 0014
Unidade de Xenética, Departamento de Anatomia Patolóxica e Ciencias Forenses, and Instituto de Ciencias Forenses, Grupo de Medicina Xenómica (GMX), Facultade de Medicina, Universidade de Santiago de Compostela, Galicia, Spain 15872
Addis Ababa University and Center of Human Genetic Diversity, P.O. Box 1176, Addis Ababa, Ethiopia
Institute of Bioorganic Chemistry Academy of Sciences Republic of Uzbekistan, Tashkent , Uzbekistan 100125
Department of Genetics and Cytology, V.N. Karazin Kharkiv National University, Kharkiv Ukraine 61077
State Office for Cultural Heritage Management Baden-Württemberg, Osteology, Konstanz, Germany D-78467
Lebanese American University, School of Medicine, Beirut, Lebanon 13-5053
Harvard School of Public Health, Boston, USA 02115
Department of Medical Biology, University of Split, School of Medicine, Split, Croatia 21000
Research Department of Genetics, Evolution and Environment, University College London, London, United Kingdom WC1E 6BT
Department of Biology and Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, USA 19104
CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India 500 007
Present address: Banaras Hindu University, Varanasi, India 221 005
Estonian Academy of Sciences, Tallinn, Estonia 10130
Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain 08003
Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA 98195
Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA 02115
Analysis of ancient DNA can reveal historical events that are difficult to discern through study of present-day individuals. To investigate European population history around the time of the agricultural transition, we sequenced complete genomes from a ~7,500 year old early farmer from the Linearbandkeramik (LBK) culture from Stuttgart in Germany and an ~8,000 year old hunter-gatherer from the Loschbour rock shelter in Luxembourg. We also generated data from seven ~8,000 year old hunter-gatherers from Motala in Sweden. We compared these genomes and published ancient DNA to new data from 2,196 samples from 185 diverse populations to show that at least three ancestral groups contributed to present-day Europeans. The first are Ancient North Eurasians (ANE), who are more closely related to Upper Paleolithic Siberians than to any present-day population. The second are West European Hunter-Gatherers (WHG), related to the Loschbour individual, who contributed to all Europeans but not to Near Easterners. The third are Early European Farmers (EEF), related to the Stuttgart individual, who were mainly of Near Eastern origin but also harbored WHG-related ancestry. We model the deep relationships of these populations and show that about ~44% of the ancestry of EEF derived from a basal Eurasian lineage that split prior to the separation of other non-Africans.

Ancient DNA studies have demonstrated that migration played a major role in the introduction of agriculture to Europe, as early European farmers were genetically distinct from ancient European hunter-gatherers\(^1,2\) and closer to present-day Near Easterners\(^2,3\). Europeans today, however, are genetically intermediate, which has led to attempts to model them as a mixture of those two ancestral populations\(^2\). A two-way mixture model is difficult to reconcile, however, with the fact that nearly all present-day Europeans also have ancestry from a third source: an Ancient North Eurasian (ANE) population\(^4,5\) that also contributed ancestry to Native Americans\(^6\).

To clarify the population transformations that accompanied the agricultural transition in Europe, we sequenced the genomes of nine ancient European individuals (Fig. 1A; Extended Data Fig. 1). We sequenced to 19-fold coverage the genome of “Stuttgart”, a ~7,500 year old individual found in Stuttgart in southern Germany who was buried in the context of artifacts from the first widespread Neolithic farming culture of central and northern Europe, the Linearbandkeramik (LBK). We sequenced to 22-fold the genome of “Loschbour”, an ~8,000 year old individual
found in the Loschbour rock shelter in Heffingen Luxembourg, from a skeleton that was discovered in the context of Mesolithic hunter-gatherer artifacts (SI1; SI2). We also sequenced DNA from seven ~8,000 year old remains from Mesolithic hunter-gatherers from the Motala site in southern Sweden, with the highest coverage individual (Motala12) at 2.4-fold. We mapped the sequences to the human reference genome (hg19), and for the two high coverage individuals (Stuttgart and Loschbour) inferred genotypes.

A central challenge in ancient DNA research is to distinguish authentic sequences from contamination. In initial sequencing libraries prepared from all nine individuals, the rate of C→T and G→A mismatches to the human genome at the ends of the DNA molecules was >20% compared with <1% for other nucleotides, suggesting authentic ancient DNA. We inferred a mitochondrial DNA (mtDNA) consensus for each sample, and based on the number of reads that differed from the consensus, estimated contamination levels of 0.3% for Loschbour, 0.4% for Stuttgart, and 0.01% - 5% for the Motala individuals. Stuttgart belonged to mtDNA haplogroup T2, typical of Neolithic Europeans, while Loschbour and all Motala individuals belonged to haplogroups U5 and U2, typical of pre-agricultural Europeans. Based on the ratio of sequences aligning to chromosomes X and Y, we infer that Stuttgart was female while Loschbour and five Motala individuals were male. Loschbour and four Motala males belonged to Y-chromosome haplogroup I, showing that this was a predominant haplogroup in pre-agricultural northern Europeans.

To generate large amounts of data, we built sequencing libraries using the enzyme uracil DNA glycosylase, which decreases the rate of C→T and G→A errors due to ancient DNA damage. After correcting for genotyping error, we estimate that heterozygosity (the number of differences per nucleotide between an individual’s two chromosomes) is 0.00074 for Stuttgart, at the high end of present-day European. Heterozygosity is 0.00048 for Loschbour, lower than in all other present-day humans we analyzed. Combined with the higher proportion of deleterious heterozygous observed in Loschbour compared with Stuttgart or present-day humans, this finding is consistent with the ancestors of Loschbour having experienced small population sizes since separation from the ancestors of the other samples. By analyzing sites known to affect phenotype, we inferred that neither Stuttgart nor Loschbour could digest milk.
into adulthood, that both had a >99% probability of dark hair, that Loschbour probably had
darker skin than Stuttgart, and that Loschbour had a >50% probability of blue eyes while
Stuttgart had a >99% probability of brown eyes (SI7). The AMY1 gene coding for salivary
amylase had 6, 13, and 16 copies in Motala12, Loschbour and Stuttgart respectively, in the range
of present-day populations (Extended Data Fig. 2) (SI 8), suggesting that high copy counts of
AMY1 in humans are not entirely due to selection since the switch to agriculture14.

To determine how the ancient genomes relate to each other and to present-day humans, we
analyzed 2,196 individuals from 185 populations genotyped at 594,924 autosomal single
nucleotide polymorphisms (SNPs) using the Affymetrix Human Origins array5 (SI9) (Extended
Data Table 1). We identified a set of “West Eurasian” populations as those that cluster with
Europe and the Near East in an ADMIXTURE15 analysis (SI9 and Extended Data Figure 3).
Principal Component Analysis (PCA)16 of the West Eurasian individuals separates Near Eastern
and European populations along parallel south-north gradients, with a handful of mostly
Mediterranean populations in between (Fig. 1B). The gradient in the Near East stretches from the
Levant to the North Caucasus, and in Europe from Sardinia to the Baltic. This plot is
qualitatively different from previous PCAs of Europeans in which the first and second PCs have
correlated well to geography17, 18; we ascribe this to our heavy sampling of Near Eastern
populations, which causes the first PC to be more dominated by European-Near Eastern
differences. We projected onto the PCs genetic data from ancient individuals2, 19, 20, which
reveals that European hunter-gatherers like Loschbour and Motala fall outside the variation of
West Eurasians in the direction of European differentiation from the Near East. This pattern is
suggestive of present-day Europeans being admixed between ancient European hunter-gatherers
and ancient Near Easterners, an inference that we confirm below. Loschbour clusters with
~7,000 year old hunter-gatherers from Spain20, allowing us to propose a “West European Hunter-
Gatherer” (WHG) meta-population. The Motala individuals cluster with ~5,000 year old
Neolithic hunter-gatherers2 from the Pitted Ware Culture (PWC) in Sweden, suggesting a
“Scandinavian Hunter-Gatherer” (SHG) meta-population that maintained biological continuity
across the Neolithic transition. Stuttgart clusters with two early farmers—the ~5,300 year old
Tyrolean Iceman19 and a ~5,000 year old southern Swedish farmer2 from the Funnel Beaker
Culture—suggesting an “Early European Farmer” (EEF) meta-population similar to present-day
Sardinians$^{19}$. Two Upper Paleolithic Siberian samples project beyond the variation of Europeans on the second PC (Fig. 2A), suggesting that they may be derive from the Ancient North Eurasian (ANE) population previously shown to have contributed to Europeans$^{4, 5}$.

PCA is a powerful technique for measuring genetic similarity, but its interpretation in terms of history is difficult, as gradients of variation due to admixture may arise under a variety of different histories$^{21}$. To test if present-day Europeans were formed by admixture of populations related to Loschbour, Stuttgart and MA1, we analyzed $f_3(X; \text{Ref}_1, \text{Ref}_2)$ statistics which measure the correlation in nucleotide frequency differences between a test sample and two populations: $(X-\text{Ref}_1)$ and $(X-\text{Ref}_2)$. If the three populations are related by a simple tree, the statistic is expected to be positive$^5$. However, if $X$ is admixed between populations related to $\text{Ref}_1$ and $\text{Ref}_2$, the statistic can be negative and provides evidence of admixture in population $X$$^5$. For each present-day West Eurasian population, we tested all possible modern reference populations with at least 4 individuals, along with Loschbour, Stuttgart, Motala12 and MA1 (Table 1). For the majority of European populations ($n=18$) the lowest $f_3$-statistic is observed with Loschbour and a Near Eastern population as references, suggesting that many Europeans derive from a mixture between WHG and populations related to present-day Near Easterners. Only Sardinians form their lowest $f_3$-statistic with Loschbour-Stuttgart so the mixture process is unlikely to have been a simple WHG-EEF one (Table 1). Other European populations form their lowest $f_3$-statistics with MA1-Stuttgart, which we hypothesize reflects the cline of increasing relatedness to MA1 in Fig. 1B. In the Near East, no population has its lowest $f_3$-statistic with Loschbour or Motala12, but all have their lowest $f_3$-statistic with Stuttgart (Table 1), suggesting that most of the ancestry of this sample may be directly inherited from populations of the ancient Near East, while modern Near Easterners have additional influences related to Africa, North Eurasia, or South Asia (Table 1).

To determine whether a mixture of just two ancestral populations can explain the negative $f_3$-statistics we observe or whether more populations are required, we analyzed $f_4$-statistics$^{5, 22}$. We began by analyzing $f_4(X, \text{Stuttgart}; \text{Loschbour}, \text{Chimpanzee})$, which measures whether Loschbour shares more alleles with West Eurasian population $X$ or with Stuttgart (Extended Data Fig. 4). This statistic is positive for nearly all Europeans showing that Stuttgart has less WHG ancestry than present-day Europeans. However, it is negative for all Near Easterners, suggesting
that the ancestors of Stuttgart were not unmixed migrants from the Near East\textsuperscript{1,2,10} (Extended Data Table 1), consistent with the clustering of Stuttgart with Europeans in the PCA of Fig. 1B. We replicated this signal in subsets of SNPs that are uniformly ascertained (Extended Data Table 2). In SI10, we estimate that the proportion of Near Eastern ancestry in Stuttgart is definitely less than 100\% and possibly as little as 61\%. Further analyses of $f_4$-statistics, however, show patterns that cannot all be explained by a history of Loschbour-related mixture. For example, the statistic $f_4(X, \text{Stuttgart}; \text{MA1}, \text{Chimpanzee})$ has a qualitatively different geographic distribution than the same statistic replacing MA1 with Loschbour, in that it is positive in both Europeans and Near Easterners whereas the latter is positive only in Europeans (Extended Data Tables 1 and 2). This and related statistics are correlated to a statistic previously shown to document a signal of ANE-related admixture into Europe\textsuperscript{4,5} (Extended Data Fig. 5), indicating that these $f_4$-statistics are reflecting ANE admixture rather than WHG admixture. Extended Data Fig. 6 visually illustrates the different admixture patterns by plotting onto a map of West Eurasia $f_4$-statistics that reflect the degree of allele sharing of each West Eurasian population with different pairs of ancient populations. We formally tested whether the $f_4$-statistic patterns are reflecting a history of more than one historical admixture event by using a method that tests the consistency of a matrix of $f_4$-statistics with descent from a specified number of ancestral populations\textsuperscript{23}. We reject the scenario of most European population descending from a mixture of just two populations (P<10\textsuperscript{-12}), but find that a scenario in which most European present-day populations descend from as few as three ancestral population is consistent with the data to the limits of our resolution (SI11).

Motivated by these observations, we modeled Europeans as a three-way mixture of ANE (of which MA1 is a member), WHG (Loschbour), and EEF (Stuttgart). To test the consistency of this model with our data, we used the ADMIXTUREGRAPH software\textsuperscript{22}, which fits a tree with discrete admixture events and reports $f$-statistics that differ by more than three standard errors between the estimated and fitted values (SI12). Our model-building was motivated by three observations (SI12): (1) Eastern non-Africans (Oceanians, East Asians, Native Americans, and Onge, indigenous Andaman islanders\textsuperscript{24}) are genetically closer to ancient Eurasian hunter-gatherers (Loschbour, Motala12 and MA1) than to Stuttgart; (2) Every eastern non-African population except Native Americans is genetically equally close to Loschbour, Motala12, and MA1, but Native Americans are genetically closer to MA1 than to European hunter-gatherers\textsuperscript{6};
and (3) All three hunter-gatherers and Stuttgart are genetically closer to Native Americans than to other eastern non-Africans. We jointly fit models to data from Loschbour, Stuttgart, MA1, Karitiana and Onge (SI12), and found that there was a unique model with two admixture events that fit the data; models with one or zero admixture events could all be rejected (SI12). One of the inferred admixture events is the ANE gene flow into both Europe⁶ and the Americas⁶ that has previously been documented. The successful model (Fig. 2A) also suggests 44 ± 10% “Basal Eurasian” admixture into the ancestors of Stuttgart: gene flow into their Near Eastern ancestors from a lineage that diverged prior to the separation of the ancestors of Loschbour and Onge. Such a scenario, while never suggested previously, is plausible given the early presence of modern humans in the Levant²⁵, African-related tools made by modern humans in Arabia²⁶,²⁷, and the geographic opportunity for continuous gene flow between the Near East and Africa²⁸.

Our fitted model (Fig. 2A) allows us to estimate fractions of ANE/WHG/EEF ancestry for each European population (SI12). To explore the robustness of these estimates, we developed an independent method for estimating mixture that only assumes that MA1 is a representative of ANE, Loschbour of WHG, and Stuttgart of EEF. Specifically, we studied $f_4$-statistics of the form $f_4(X, \text{Stuttgart}; \text{Outgroup}_1, \text{Outgroup}_2)$, measuring the correlation in allele frequency difference between $X$ and Stuttgart, and a pair of outgroups with no recent shared history with Europeans. We chose divergent outgroups (SI13) that are differentially related to ANE, WHG and EEF, and then expressed the $f_4$-statistics for each European population as a linear combination of $f_4(\text{Loschbour, Stuttgart}; \text{Outgroup}_1, \text{Outgroup}_2)$ and $f_4(\text{MA1, Stuttgart}; \text{Outgroup}_1, \text{Outgroup}_2)$, fitting the mixture coefficients that minimize the difference between expected and observed $f_4$-statistics. The mixture coefficients agree between this method and the ADMIXTUREGRAPH modeling, increasing our confidence in both analyses (Extended Data Table 3, SI13).

Our estimates of mixture proportions (Fig. 2B and Extended Table 3) indicate that EEF ancestry in Europe today ranges from as little as around 30% in the Baltic to as high as around 90% in the Mediterranean (a previous study² inferred 11% in Russians to 95% in Sardinians, but fit a two-population mixture model). The north-south gradient is also consistent with patterns of identity-by-descent (IBD) sharing²⁹, in which Loschbour shares more segments with northern Europeans and Stuttgart with southern Europeans (SI14). We infer that southern Europeans received their
European hunter-gatherer ancestry mostly via EEF, while Northern Europeans acquired up to 50% additional WHG-related ancestry. Europeans also have ANE ancestry (up to ~20%), which is widespread across Europe, but quantitative less as the WHG/(WHG+ANE) ratio is ~0.6-0.8 for most Europeans (SI12). The history behind the ANE ancestry in West Eurasia is not simple, as the Near East has little or no WHG ancestry but substantial levels of ANE ancestry there especially in the North Caucasus (SI12; Fig.1B; Fig. 2). Loschbour and Stuttgart had little or no ANE ancestry, indicating that it was not as pervasive in central Europe around the time of the agricultural transition as it is today. (By implication ANE ancestry was also not present in the ancient Near East; since Stuttgart which has substantial Near Eastern ancestry lacks it.) However, ANE ancestry was already present in at least some Europeans (Scandinavian hunter-gatherers) by ~8,000 years ago, since MA1 shares more alleles with Motala12 than Loschbour: \( f_4(\text{Motala12}; \text{Loschbour}; \text{MA1}, \text{Mbuti}) = 0.003 \) (Z=5.2 standard errors from zero) (SI12). While SHG may have contributed ANE ancestry to modern Europeans, it cannot have been the only population that did so, as no European population has its lower \( f_3 \)-statistic with it in Table 1, and few populations fit a model of EEF-SHG admixture (SI12).

While our three-way mixture model fits the data for most European populations, two sets of populations are poor fits. First, Sicilians, Maltese, and Ashkenazi Jews have EEF estimates beyond the 0-100% interval (SI13) and they cannot be jointly fit with other Europeans in the (SI12). These populations may have more Near Eastern ancestry than can be explained via EEF admixture (SI13), an inference that is also suggested by the fact that they fall in the gap between European and Near Eastern populations in the PCA of Fig. 1B. Second, we observe that Finns, Mordovians, Russians, Chuvash, and Saami from northeastern Europe do not fit our model (SI12; Extended Data Table 3). To better understand this, for each West Eurasian population in turn we plotted \( f_4(X, \text{Bedouin2}; \text{Han}, \text{Mbuti}) \) against \( f_4(X, \text{Bedouin2}; \text{MA1}, \text{Mbuti}) \), using statistics that measure the degree of a European population’s allele sharing with Han Chinese or MA1 (Extended Data Fig. 7). Europeans fall along a line of slope >1 in the plot of these two statistics. However, northeastern Europeans fall away from this line in the direction of Han. This is consistent with Siberian gene flow into some northeastern Europeans after the initial ANE admixture, and may be related to the fact that Y-chromosome haplogroup \( N^{30,31} \) is shared between Siberian and northeastern Europeans\(^{32,33} \) but not with western Europeans. There may in
fact be multiple layers of Siberian gene flow into northeastern Europe after the initial ANE gene flow, as our analyses reported in SI 12 show that some Mordovians, Russians and Chuvash have Siberian-related admixture that is significantly more recent than that in Finns (SI12).

This study raises two questions that are important to address in future research. A first is where the EEF picked up their WHG ancestry. Southeastern Europe is a candidate as it lies along the geographic path from Anatolia into central Europe, and hence it should be a priority to study ancient samples from this region. A second question is when and where ANE ancestors admixed with the ancestors of most present-day Europeans. Based on discontinuity in mtDNA haplogroup frequencies in Central Europe, this may have occurred during the Late Neolithic or early Bronze Age \(\text{~5,500-4,000 years ago}\)\(^{35}\). A central aim for future work should be to collect transects of ancient Europeans through time and space to illuminate the history of these transformations.
Supplementary Information is linked to the online version of the paper. The fully public version of the Human Origins dataset can be found at http://genetics.med.harvard.edu/reichlab/Reich_Lab/Datasets.html, whereas a version with additional samples that require users to sign a letter indicating that they will abide by specified usage conditions is available on request from DR.

Acknowledgments
We are grateful to Cynthia Beall, Neil Bradman, Mark Shriver, Amha Gebremedhin, Sena Karachanak-Yankova, Damian Labuda, Theologos Loukidis and Anna Di Rienzo for sharing DNA samples; to Detlef Weigel, Christa Lanz, Verena Schünemann, Peter Bauer and Olaf Riess for support and access to DNA sequencing facilities; to Nadin Rohland for sample handling; to Arti Tandon for bioinformatic support; to Philip Johnson for advice on contamination estimation; and to Pontus Skoglund for sharing the graphics software used to generate Extended Data Figure 6. We thank all the volunteers who donated DNA, and the staff of the Unità Operativa Complessa di Medicina Trasfusionale, Azienda Ospedaliera Umberto I, Siracusa, Italy for assistance in sample collection. JK is grateful for support from DFG grant # KR 4015/1-1, the Carl-Zeiss Foundation and the Baden Württemberg Foundation. SP acknowledges support from the Presidential Innovation Fund of the Max Planck Society. EB and OB were supported by RFBR grants 13-06-00670, 13-04-01711, 13-04-90420 and by the Molecular and Cell Biology Program of the Presidium, Russian Academy of Sciences. OB was supported by GFFI grant 53.4/071. BM was supported by grants OTKA 73430 and 103983. The Lithuanian sample collection was supported by the LITGEN project (VP1-3.1-ŠMM-07-K-01-013), funded by the European Social Fund under the Global Grant Measure. AS was supported by Spanish grants SAF2008-02971 and EM 2012/045. OU was supported by Ukrainian SFFS grant F53.4/071. SAT was supported by NIH Pioneer Award 8DP1ES022577-04 and NSF HOMINID Award BCS-0827436. KT was supported by an Indian CSIR Network Project (GENESIS: BSC0121). LS was supported by an Indian CSIR Bhatnagar Fellowship. RV, MM, JP and EM were supported by the European Union Regional Development Fund through the Centre of Excellence in Genomics to the Estonian Biocentre and University of Tartu and by a Estonian Basic Research grant SF0270177As08. MM was additionally supported by Estonian Science Foundation grant #8973. JGS and MS were supported by NIH grant GM40282. PHS and EEE were supported by
NIH grants HG004120 and HG002385. DR and NP were supported by NSF HOMINID grant BCS-1032255 and NIH grant GM100233. DR and EEE are Howard Hughes Medical Institute investigators.

**Author contributions**

EEE, JBu, MS, SP, JKe, DR and JKr supervised the study. IL, NP, AM, GR, SM, PHS, JGS, SC, KK, QF, CdF, KP, WH, MMey, and DR analyzed genetic data. FH, EF, DD, MF, J-MG, JW, AC and JKr obtained archaeological material. AM, CE, RB, KB, SS, CP and JKr processed ancient DNA. IL, NP, SN, GA, HAB, EB, OB, HB-A, JBe, FBe, FBr, GBJB, FC, MC, DECC, LD, GvD, SD, SAF, IGR, MG, MH, BH, TH, UH, ARJ, RKi, EK, TK, VK, RKh, AK, LL, SL, RWM, BM, EM, JM, TN, LO, JP, FP, OLP, VR, IR, RR, HS, AS, EBS, AT, DT, ST, IU, OU, MV, PZ, LY, TZ, CC, MGT, SAT, LS, KT, RV, DC, RS, MMet, SP and DR assembled the genotyping dataset. IL, DR and JKr wrote the manuscript with help from all co-authors.

**Author information**

The authors declare no competing financial interests except for JM who is an employee of 23andMe.
Table 1: Lowest $f_3$-statistics for each West Eurasian population

<table>
<thead>
<tr>
<th>$Ref_1$</th>
<th>$Ref_2$</th>
<th>Target pop. for which these two reference pops. give the most negative $f_3(X; Ref_1, Ref_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loschbour</td>
<td>Stuttgart</td>
<td>Sardinian</td>
</tr>
<tr>
<td>Loschbour</td>
<td>Near East (Abkhasian, Armenian, Georgian, Iraqi_Jew)</td>
<td>Estonian, Finnish, Icelandic, Lithuanian, Mordovian, Russian, Orcadian, Belorussian, Czech, Norwegian, Ukrainian, Basque, English, French_South, Spain_North, Scottish, Spanish</td>
</tr>
<tr>
<td>Stuttgart</td>
<td>North Eurasia (MA, Piapoco)</td>
<td>Albanian, Ashkenazi_Jew, Bergamo, Bulgarian, Croatian, French, Greek, Hungarian, Maltese, Sicilian, Tuscan, Abkhasian, Chechen, Cypriot, Druze, Lezgin, Turkish_Jew, Adygei, Balkar, Iranian, Kumyk, North_Ossetian, Turkish</td>
</tr>
<tr>
<td>Stuttgart</td>
<td>South Asian (Gujarati, Vishwabrahmin)</td>
<td>Armenian, Georgian, Georgian_Jew, Iranian_Jew, Iraqi_Jew</td>
</tr>
</tbody>
</table>

Note: We group populations into five categories of similar $Ref_1$ and $Ref_2$, using a capital letter superscript to indicate the $Ref_2$ population in each pair. The Z-scores for populations are < -4 except where marked in gray. Extended Table 1 gives the quantitative values of each $f_3$-statistic.
**Figure Legends**

**Figure 1: Map of West Eurasian populations and Principal Component Analysis.** (a) Locations of ancient and present-day samples analyzed, with color coding matching the PCA. We show all sampling locations for each population, which results in multiple points for some populations (e.g. Spain). (b) PCA on all present-day West Eurasians, with the ancient and selected eastern non-Africans projected. European hunter-gatherers fall beyond modern Europe in the direction of European differentiation from the Near East. Stuttgart clusters with other Neolithic Europeans and present-day Sardinians. MA1 falls outside the variation of modern day West Eurasians in the direction of southern-northern differentiation along dimension 2 and between the European and Near Eastern clines along dimension 1.

**Figure 2: Modeling of West Eurasian population history.** (a) A three-way mixture model that is a statistical fit to the data for many European populations, ancient DNA samples, and non-European populations. Present-day samples are colored in blue, ancient samples in red, and reconstructed ancestral populations in green. Solid lines represent descent without mixture, and dashed lines represent admixture events. For the two mixture events relating the highly divergent ancestral populations, we print estimates for the mixture proportions as well as one standard error. (b) We plot the proportions of ancestry from each of three inferred ancestral populations (EEF, ANE and WHG) as inferred from the model-based analysis.
Figure 1

A

B

Modern West Eurasians
- Armenian
- Iranian
- Turkish
- Albanian
- Bergamo
- Bulgarian
- Syriot
- Greek
- Maltese
- Sicilian
- Tuscan
- English
- Abkhazian
- French
- Icelandic
- Norwegian
- Orcadian
- Scottish
- Bedouin-A
- Bedouin-B
- Jordanian
- Palestinian
- Saudi
- Syrian
- Abkhazian

Adygei
- Bakkar
- Chechen
- Georgian
- Kumyk
- Laccin
- North Ossetian
- Ashkenazi_Jew
- Georgian_Jew
- Iranian_Jew
- Iraqi_Jew
- Libyan_Jew

Moroccan_Jew
- Tunisian_Jew
- Turkish_Jew
- Yemenite_Jew
- Basque
- French_South
- Pais Vasco
- Spanish
- Druze
- Lebanese
- Belorussian
- Russian

Czech
- Estonian
- Hungarian
- Lithuanian
- Ukrainian
- Canary_Islanders
- Sardian
- Spanish
- Finnish
- Moldovan

Eastern non-Africans (projected)
- Ami
- Bougainville
- Atyal
- Papuan
- Han
- Onge
- Naxi
- Kartiana

Ancients (projected)
- Stuttgart
- Skoglund_farmer
- Iceman
- Loschbour
- LaBrana1
- LaBrana2
- Motla12
- Motla_merge
- Skoglund_mergeHG
- MA1
- AG2

Ancient North Eurasians
Scandinavian Hunter-Gatherers
Western European Hunter-Gatherers
Early European Farmers
Figure 2

A

B

ANAE

WHG

EEF

• Albanian  ■ French  ▲ French_South  ● Estonian
• Bergamo  ▲ Icelandic  ▲ Pais_Vasco  ❌ Hungarian
• Bulgarian  ▲ Norwegian  ● Spanish  ● Lithuanian
• Greek  ❌ Orcadian  ● Belorussian  + Ukrainian
• Tuscan  ● Scottish  ■ Croatian  ▲ Czech
• English  ● Basque  ▲ Sardinian
References


Ville N. Pimenoff, David Comas, Jukka U. Palo, Galina Vershubsky, Andrew Kozlov, and Antti Sajantila, 'Northwest Siberian Khanty and Mansi in the Junction of West and East Eurasian Gene Pools as Revealed by Uniparental Markers', *Eur J Hum Genet*, 16 (2008), 1254-64.
