bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Iosif Lazaridis[1,2], Nick Patterson[1,2], Alissa Mittnik[3], Gabriel Renaud[4], Swapan Mallick[1,2], Peter H. Sudmant[5], Joshua G. Schraiber[6], Sergi Castellano[4], Karola Kirsanow[7], Christos Economou[8], Ruth Bollongino[7], Qiaomei Fu[4], Kirsten I. Bos[3], Susanne Nordenfelt[1,2], Cesare de Filippo[4], Kay Prüfer[4], Susanna Sawyer[4], Cosimo Posth[3], Wolfgang Haak[9], Fredrik Hallgren[10], Elin Fornander[10], George Ayodo[11], Hamza A. Babiker[12,13], Elena Balanovska[14], Oleg Balanovsky[14,15], Haim Ben-Ami[16], Judit Bene[17], Fouad Berrada[18], Francesca Brisighelli[19], George Busby[20,21], Francesco Cali[22], Mikhail Churnosov[23], David E. C. Cole[24], Larissa Damba[25], Dominique Delsate[26,27], George van Driem[28], Stanislav Dryomov[29,30], Sardana A. Fedorova[31], Michael Francken[32], Irene Gallego Romero[33], Marina Gubina[34], Jean-Michel Guinet[26], Michael Hammer[35], Brenna Henn[36], Tor Hervig[37], Ugur Hodoglugil[38], Aashish R. Jha[33], Rick Kittles[39], Elza Khusnutdinova[40,41], Toomas Kivisild[42], Vaidutis Kučinskas[43], Rita Khusainova[40,41], Alena Kushniarevich[44], Leila Laredj[45], Sergey Litvinov[40,41,44], Robert W. Mahley[46], Béla Melegh[17], Ene Metspalu[47], Joanna Mountain[48], Thomas Nyambo[49], Ludmila Osipova[34], Jüri Parik[47], Fedor Platanov[50], Olga Posukh[34], Valentino Romano[51], Igor Rudan[52], Ruslan Ruizbakiev[53], Hovhannes Sahakyan[44,54], Antonio Salas[55], Elena B. Starikovskaya[29,30], Ayele Tarekegn[56], Draga Toncheva[Error! Bookmark not defined.], Shahlo Turdikulova[57], Ingrida Uktveryte[43], Olga Utevska[58], Mikhail Voevoda[25,34], Joachim Wahl[59], Pierre Zalloua[60,61], Levon Yepiskoposyan[54], Tatijana Zemunik[62], Alan Cooper[9], Cristian Capelli[21], Mark G. Thomas[63], Sarah A. Tishkoff[64], Lalji Singh[65,66], Kumarasamy Thangaraj[65], Richard Villems[44,47,67], David Comas[68], Rem Sukernik[29,30], Mait Metspalu[44], Matthias Meyer[4], Evan E. Eichler[5,69] , Joachim Burger[7], Montgomery Slatkin[6], Svante Pääbo[4], Janet Kelso[4], David Reich[1,2,70,†] and Johannes Krause[3,†]

† Co-senior authors.

To whom correspondence should be addressed: David Reich (reich@genetics.med.harvard.edu) or Johannes Krause (johannes.krause@uni-tuebingen.de)

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[1] Department of Genetics, Harvard Medical School, Boston, MA, USA 02115
[2] Broad Institute of Harvard and MIT, Cambridge, MA, USA 02142
[3] Institute for Archaeological Sciences, University of Tübingen, Rümelinstrasse 23, Tübingen, Germany 72072
[4] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany 04103
[5] Department of Genome Sciences, University of Washington, Seattle, WA, USA 98195
[6] Department of Integrative Biology, University of California, Berkeley, CA 94720-3140
[7] Johannes Gutenberg University Mainz, Institute of Anthropology, Mainz, Germany D-55128
[8] Archaeological Research Laboratory, Stockholm University, Stockholm, Sweden 114 18
[9] Australian Centre for Ancient DNA, School of Earth and Environmental Sciences, University of Adelaide, Adelaide, South Australia, Australia SA 5005
[10] The Cultural Heritage Foundation, Västerås, Sweden 722 12
[11] Center for Global Health and Child Development, Kisumu, Kenya 40100
[12] Institutes of Evolution, Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, Edinburgh, UK EH9 3JT
[13] Biochemistry Department, Faculty of Medicine, Sultan Qaboos University, Alkhod, Muscat, Oman 123
[14] Research Centre for Medical Genetics, Moscow , Russia 115478
[15] Vavilov Institute for General Genetics, Moscow, Russia 119991
[16] Rambam Health Care Campus, Haifa 31096, Israel
[17] Department of Medical Genetics, and Szentagothai Research Center, University of Pécs, Pécs, Hungary 7624
[18] Al Akhawayn University in Ifrane (AUI), School of Science and Engineering, Ifrane, Morocco 53000
[19] Forensic Genetics Laboratory, Institute of Legal Medicine, Università Cattolica del Sacro Cuore, Rome, Italy 00168
[20] Department of Zoology, University of Oxford, Oxford, UK OX1 3PS
[21] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK OX3 7BN
[22] Laboratorio di Genetica Molecolare, IRCCS Associazione Oasi Maria SS, Troina Italy 94018
[23] Belgorod State University, Belgorod, Russia
[24] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada M5G 1L5
[25] Institute of Internal Medicine, Siberian Branch of Russian Academy of Medical Sciences, Novosibirsk, Russia 630089
[26] Luxembourg Natural History Museum, Luxembourg L-2160
[27] Centre National de Recherche Archéologique auprès du Musée National d'Histoire et d'Art, Luxembourg
[28] Institute of Linguistics, University of Bern, Bern, Switzerland CH-3012
[29] Laboratory of Human Molecular Genetics, Institute of Molecular and Cellular Biology, Russian Academy of Science, Siberian Branch, Novosibirsk, Russia 630090
[30] Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch, Novosibirsk, Russia 630090.
[31] Yakut Research Center of Complex Medical Problems and North-Eastern Federal University, Yakutsk, Russia, 677010
[32] Department of Paleoanthropology, Senckenberg Center for Human Evolution and Paleoenvironment, University of Tübingen, Tübingen, Germany D-72070
[33] Department of Human Genetics, University of Chicago, Chicago, IL USA 60637
[34] The Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia 630090
[35] ARL Division of Biotechnology, University of Arizona, Tucson, AZ, USA 85721
[36] Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA 11794
[37] Department of Clinical Science, University of Bergen, Bergen, Norway 5021

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[38] NextBio, part of Illumina, Santa Clara, CA, USA 95050
[39] College of Medicine, University of Illinois at Chicago, Chicago, IL, USA 60607
[40] Institute of Biochemistry and Genetics, Ufa Research Centre, Russian Academy of Sciences, Ufa, Russia 450054
[41] Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia 450074
[42] Division of Biological Anthropology, University of Cambridge, Cambridge, United Kingdom CB2 1QH
[43] Department of Human and Medical Genetics, Vilnius University, Vilnius, Lithuania LT-08661
[44] Estonian Biocentre, Evolutionary Biology group, Tartu, Estonia 51010
[45] Translational Medicine and Neurogenetics, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France 67404
[46] Robert W. Mahley, Gladstone Institutes, San Francisco, CA 94158
[47] Department of Evolutionary Biology, University of Tartu, Tartu, Estonia 51010
[48] Research Department, 23andMe, Inc. Mountain View, CA, USA 94043
[49] Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania
[50] Research Institute of Health, North-Eastern Federal University, Yakutsk, Russia 677000
[51] Dipartimento di Fisica e Chimica, Università di Palermo, Palermo, Italy 90128
[52] Centre for Population Health Sciences, The University of Edinburgh Medical School, Edinburgh, Scotland, United Kingdom EH8 9AG
[53] Deceased: formerly of the Institute of Immunology, Academy of Science, Tashkent, Uzbekistan 70000
[54] Laboratory of Ethnogenomics, Institute of Molecular Biology, National Academy of Sciences of Armenia, Yerevan, Armenia 0014
[55] Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses, and Instituto de Ciencias Forenses, Grupo de Medicina Xenómica (GMX), Facultade de Medicina, Universidade de Santiago de Compostela, Galcia, Spain 15872
[56] Addis Ababa University and Center of Human Genetic Diversity, P.O. Box 1176, Addis Ababa, Ethiopia
[57] Institute of Bioorganic Chemistry Academy of Sciences Republic of Uzbekistan, Tashkent , Uzbekistan 100125
[58] Department of Genetics and Cytology, V.N. Karazin Kharkiv National University, Kharkiv Ukraine 61077
[59] State Office for Cultural Heritage Management Baden-Württemberg, Osteology, Konstanz, Germany D-78467
[60] Lebanese American University, School of Medicine, Beirut, Lebanon 13-5053
[61] Harvard School of Public Health, Boston, USA 02115
[62] Department of Medical Biology, University of Split, School of Medicine, Split, Croatia 21000
[63] Research Department of Genetics, Evolution and Environment, University College London, London, United Kingdom WC1E 6BT
[64] Department of Biology and Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, USA 19104
[65] CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India 500 007
[66] Present address: Banaras Hindu University, Varanasi, India 221 005
[67] Estonian Academy of Sciences, Tallinn, Estonia 10130
[68] Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain 08003
[69] Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA 98195
[70] Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA 02115

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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[7] (SI2).

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[8, 9] (SI3). 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 (SI3). Stuttgart belonged to mtDNA haplogroup T2, typical of Neolithic Europeans[10], while Loschbour and all Motala individuals belonged to haplogroups U5 and U2, typical of pre-agricultural Europeans[1, 8] (SI4). 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[11] (SI5). Loschbour and four Motala males belonged to Y-chromosome haplogroup I, showing that this was a predominant haplogroup in pre-agricultural northern Europeans[12, 13] (SI5).

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 (SI3). 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 (SI2). 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 (SI6), 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

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 agriculture[14].

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 array[5] (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 ADMIXTURE[15] 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 geography[17, 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 individuals[2, 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 Spain[20], allowing us to propose a "West European Hunter-Gatherer" (WHG) meta-population. The Motala individuals cluster with ~5,000 year old Neolithic hunter-gatherers[2] 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 Iceman[19] and a ~5,000 year old southern Swedish farmer[2] from the Funnel Beaker Culture—suggesting an "Early European Farmer" (EEF) meta-population similar to present-day

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 $f_3(X; Ref_1, Ref_2)$ 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 Ref_1 and 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 $f_4(X, Stuttgart; Loschbour, Chimpanzee)$, 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

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that the ancestors of Stuttgart were not unmixed migrants from the Near East[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, Stuttgart; MA1, 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[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[23]. We reject the scenario of most European population descending from a mixture of just two populations ($P<10^{-12}$), 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[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[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[6];

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 Europe[6] and the Americas[6] 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 Levant[25], African-related tools made by modern humans in Arabia[26, 27], and the geographic opportunity for continuous gene flow between the Near East and Africa[28].

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 $f_4(X, Stuttgart; Outgroup_1, Outgroup_2)$, 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 $f_4(Loschbour, Stuttgart; Outgroup_1, Outgroup_2)$ and $f_4(MA1, Stuttgart; Outgroup_1, Outgroup_2)$, 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 study[2] 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) sharing[29], in which Loschbour shares more segments with northern Europeans and Stuttgart with southern Europeans (SI14). We infer that southern Europeans received their

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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*(Motala12; Loschbour; MA1, 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, Bedouin2; Han, Mbuti)* against f_4*(X, Bedouin2; MA1, 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

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 ~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.

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Lowest f_3-statistics for each West Eurasian population

Ref₁	***Ref₂***	**Target pop. for which these two reference pops. give the most negative $f_3(X; Ref_1, Ref_2)$**
Loschbour	**Stuttgart**	Sardinian
Loschbour	**Near East** (Abkhasian[A], Armenian[B], Georgian[C], Iraqi_Jew[D])	Estonian[A], Finnish[A], Icelandic[A], Lithuanian[A], Mordovian[A], Russian[A], Orcadian[B], Belorussian[C], Czech[C], Norwegian[C], Ukrainian[C], Basque[D], English[D], French_South[D], Spain_North[D], Scottish[D], Spanish[D]
Stuttgart	**North Eurasia** (MA1[E], Piapoco[F])	Albanian[E], Ashkenazi_Jew[E], Bergamo[E], Bulgarian[E], Croatian[E], French[E], Greek[E], Hungarian[E], Maltese[E], Sicilian[E], Tuscan[E], Abkhasian[E], Chechen[E], Cypriot[E], Druze[E], Lezgin, Turkish_Jew[E]; Adygei[F], Balkar[F], Iranian[F], Kumyk[F], North_Ossetian[F], Turkish[F]
Stuttgart	**African** (Esan[G], Gambian[H], Kgalagadi[I])	BedouinA[G], BedouinB[G], Jordanian[G], Lebanese[G], Libyan_Jew[G], Moroccan_Jew[G], Palestinian[G], Syrian[G], Yemenite_Jew[G]; Tunisian_Jew[H]; Saudi[I]
Stuttgart	**South Asian** (Gujarati3[J], Vishwabrahmin[K])	Armenian[J], Georgian[J], Georgian_Jew[J], Iranian_Jew[J], Iraqi Jew[K]

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.

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1

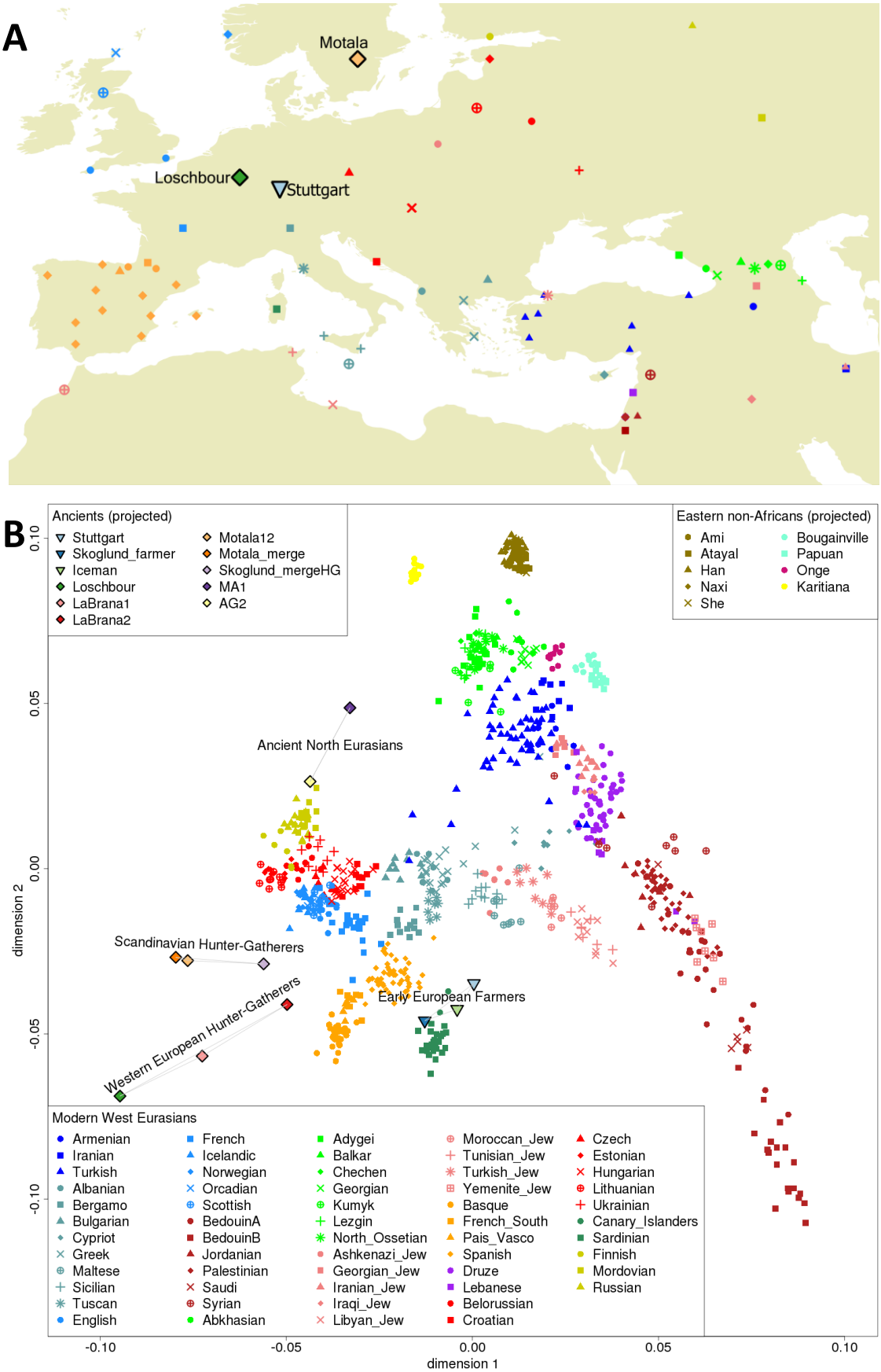
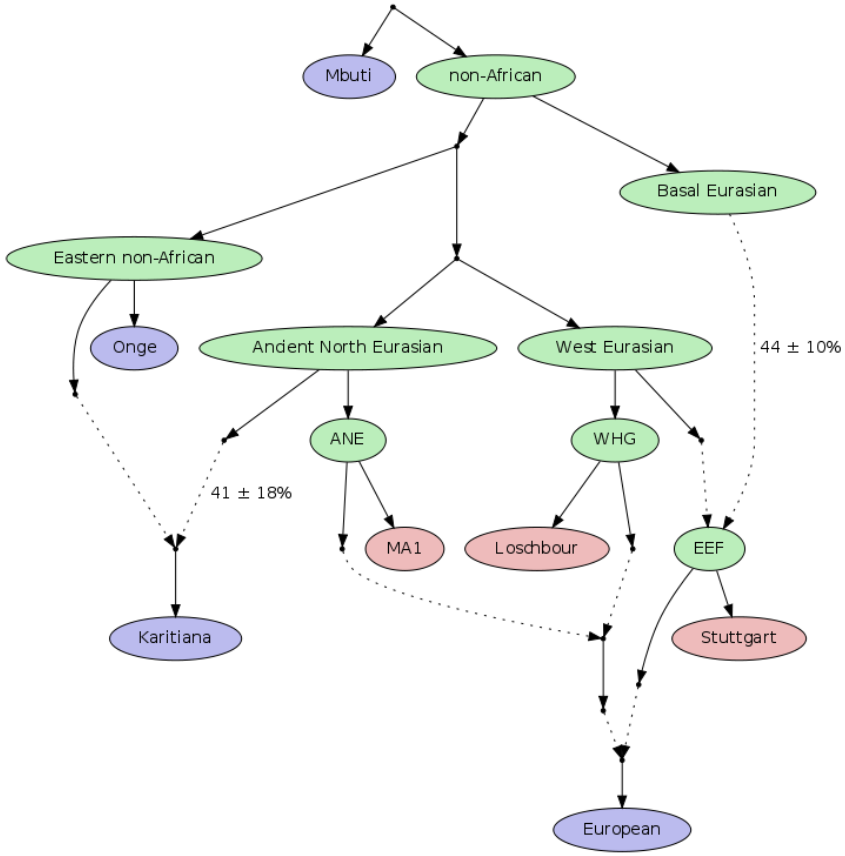
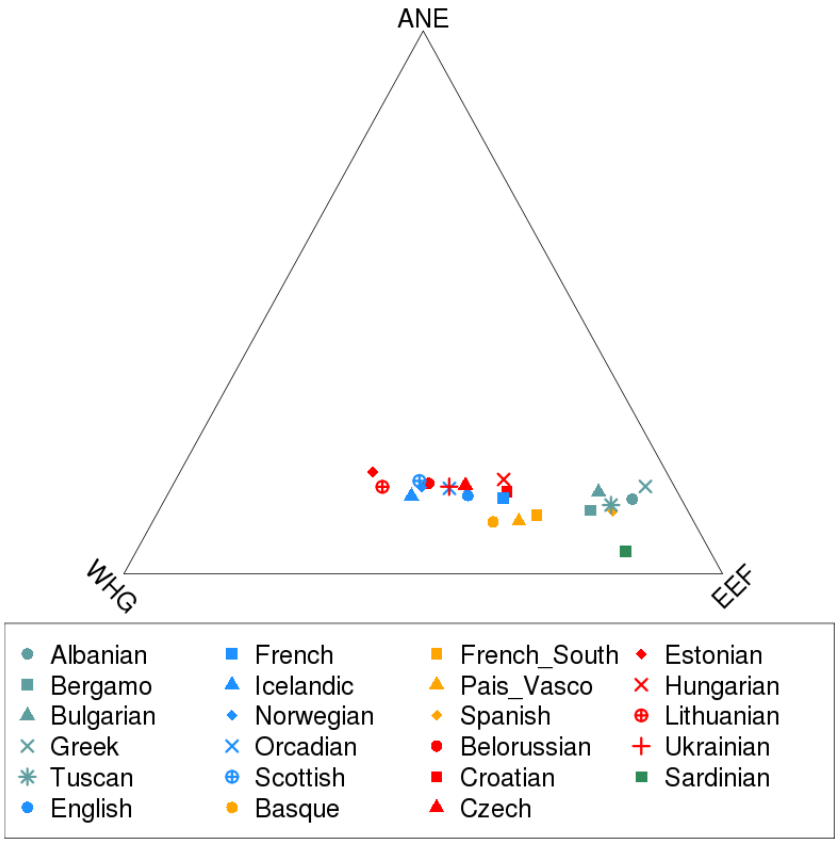


Figure 2

A


Mbuti
non-African
Basal Eurasian
Eastern non-African
Onge
Ancient North Eurasian
West Eurasian
44 ± 10%
ANE
WHG
41 ± 18%
MA1
Loschbour
EEF
Karitiana
Stuttgart
European

B


ANE
WHG
EEF
Albanian
Bergamo
Bulgarian
Greek
Tuscan
English
French
Icelandic
Norwegian
Orcadian
Scottish
Basque
French_South
Pais_Vasco
Spanish
Belorussian
Croatian
Czech
Estonian
Hungarian
Lithuanian
Ukrainian
Sardinian

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1 B. Bramanti, M. G. Thomas, W. Haak, M. Unterlaender, P. Jores, K. Tambets, I. Antanaitis-Jacobs, M. N. Haidle, R. Jankauskas, C.-J. Kind, F. Lueth, T. Terberger, J. Hiller, S. Matsumura, P. Forster, and J. Burger, 'Genetic Discontinuity between Local Hunter-Gatherers and Central Europe's First Farmers', *Science*, 326 (2009), 137-40.

2 Pontus Skoglund, Helena Malmström, Maanasa Raghavan, Jan Storå, Per Hall, Eske Willerslev, M. Thomas P. Gilbert, Anders Götherström, and Mattias Jakobsson, 'Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe', *Science*, 336 (2012), 466-69.

3 Wolfgang Haak, Oleg Balanovsky, Juan J. Sanchez, Sergey Koshel, Valery Zaporozhchenko, Christina J. Adler, Clio S. I. Der Sarkissian, Guido Brandt, Carolin Schwarz, Nicole Nicklisch, Veit Dresely, Barbara Fritsch, Elena Balanovska, Richard Villems, Harald Meller, Kurt W. Alt, Alan Cooper, and Consortium the Genographic, 'Ancient DNA from European Early Neolithic Farmers Reveals Their near Eastern Affinities', *PLoS Biol*, 8 (2010), e1000536.

4 Mark Lipson, Po-Ru Loh, Alex Levin, David Reich, Nick Patterson, and Bonnie Berger, 'Efficient Moment-Based Inference of Admixture Parameters and Sources of Gene Flow', *Molecular Biology and Evolution*, 30 (2013), 1788-802.

5 N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, and D. Reich, 'Ancient Admixture in Human History', *Genetics*, 192 (2012), 1065-93.

6 Maanasa Raghavan, Pontus Skoglund, Kelly E. Graf, Mait Metspalu, Anders Albrechtsen, Ida Moltke, Simon Rasmussen, Thomas W. Stafford Jr, Ludovic Orlando, Ene Metspalu, Monika Karmin, Kristiina Tambets, Siiri Rootsi, Reedik Magi, Paula F. Campos, Elena Balanovska, Oleg Balanovsky, Elza Khusnutdinova, Sergey Litvinov, Ludmila P. Osipova, Sardana A. Fedorova, Mikhail I. Voevoda, Michael DeGiorgio, Thomas Sicheritz-Ponten, Soren Brunak, Svetlana Demeshchenko, Toomas Kivisild, Richard Villems, Rasmus Nielsen, Mattias Jakobsson, and Eske Willerslev, 'Upper Palaeolithic Siberian Genome Reveals Dual Ancestry of Native Americans', *Nature*, advance online publication (2013).

7 A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo, 'The Genome Analysis Toolkit: A Mapreduce Framework for Analyzing Next-Generation DNA Sequencing Data', *Genome Res*, 20 (2010), 1297-303.

8 Johannes Krause, Adrian W. Briggs, Martin Kircher, Tomislav Maricic, Nicolas Zwyns, Anatoli Derevianko, and Svante Pääbo, 'A Complete Mtdna Genome of an Early Modern Human from Kostenki, Russia', *Current biology : CB*, 20 (2010), 231-36.

9 Susanna Sawyer, Johannes Krause, Katerina Guschanski, Vincent Savolainen, and Svante Pääbo, 'Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA', *PLoS ONE*, 7 (2012), e34131.

10 Wolfgang Haak, Peter Forster, Barbara Bramanti, Shuichi Matsumura, Guido Brandt, Marc Tänzer, Richard Villems, Colin Renfrew, Detlef Gronenborn, Kurt Werner Alt, and Joachim Burger, 'Ancient DNA from the First European Farmers in 7500-Year-Old Neolithic Sites', *Science*, 310 (2005), 1016-18.

11 Pontus Skoglund, Jan Storå, Anders Götherström, and Mattias Jakobsson, 'Accurate Sex Identification of Ancient Human Remains Using DNA Shotgun Sequencing', *Journal of Archaeological Science*, 40 (2013), 4477-82.

12 Siiri Rootsi, Toomas Kivisild, Giorgia Benuzzi, Hela Help, Marina Bermisheva, Ildus Kutuev, Lovorka Barać, Marijana Peričić, Oleg Balanovsky, Andrey Pshenichnov, Daniel Dion, Monica Grobei, Lev A. Zhivotovsky, Vincenza Battaglia, Alessandro Achilli, Nadia Al-Zahery, Jüri Parik, Roy King, Cengiz Cinnioğlu, Elsa Khusnutdinova, Pavao Rudan, Elena Balanovska, Wolfgang Scheffrahn, Maya Simonescu, Antonio Brehm, Rita Goncalves, Alexandra Rosa, Jean-Paul Moisan, Andre Chaventre, Vladimir Ferak, Sandor Füredi, Peter J. Oefner, Peidong Shen, Lars Beckman, Ilia Mikerezi, Rifet Terzić, Dragan Primorac, Anne Cambon-Thomsen, Astrida

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Krumina, Antonio Torroni, Peter A. Underhill, A. Silvana Santachiara-Benerecetti, Richard Villems, Chiara Magri, and Ornella Semino, 'Phylogeography of Y-Chromosome Haplogroup I Reveals Distinct Domains of Prehistoric Gene Flow in Europe', *The American Journal of Human Genetics*, 75 (2004), 128-37.

13 Pedro Soares, Alessandro Achilli, Ornella Semino, William Davies, Vincent Macaulay, Hans-Jürgen Bandelt, Antonio Torroni, and Martin B. Richards, 'The Archaeogenetics of Europe', *Current Biology*, 20 (2010), R174-R83.

14 George H. Perry, Nathaniel J. Dominy, Katrina G. Claw, Arthur S. Lee, Heike Fiegler, Richard Redon, John Werner, Fernando A. Villanea, Joanna L. Mountain, Rajeev Misra, Nigel P. Carter, Charles Lee, and Anne C. Stone, 'Diet and the Evolution of Human Amylase Gene Copy Number Variation', *Nat Genet*, 39 (2007), 1256-60.

15 D. H. Alexander, J. Novembre, and K. Lange, 'Fast Model-Based Estimation of Ancestry in Unrelated Individuals', *Genome Res*, 19 (2009), 1655-64.

16 N. Patterson, A. L. Price, and D. Reich, 'Population Structure and Eigenanalysis', *PLoS Genet*, 2 (2006), e190.

17 Oscar Lao, Timothy T. Lu, Michael Nothnagel, Olaf Junge, Sandra Freitag-Wolf, Amke Caliebe, Miroslava Balascakova, Jaume Bertranpetit, Laurence A. Bindoff, David Comas, Gunilla Holmlund, Anastasia Kouvatsi, Milan Macek, Isabelle Mollet, Walther Parson, Jukka Palo, Rafal Ploski, Antti Sajantila, Adriano Tagliabracci, Ulrik Gether, Thomas Werge, Fernando Rivadeneira, Albert Hofman, André G. Uitterlinden, Christian Gieger, Heinz-Erich Wichmann, Andreas Rüther, Stefan Schreiber, Christian Becker, Peter Nürnberg, Matthew R. Nelson, Michael Krawczak, and Manfred Kayser, 'Correlation between Genetic and Geographic Structure in Europe', *Current Biology*, 18 (2008), 1241-48.

18 John Novembre, Toby Johnson, Katarzyna Bryc, Zoltan Kutalik, Adam R. Boyko, Adam Auton, Amit Indap, Karen S. King, Sven Bergmann, Matthew R. Nelson, Matthew Stephens, and Carlos D. Bustamante, 'Genes Mirror Geography within Europe', *Nature*, 456 (2008), 98-101.

19 Andreas Keller, Angela Graefen, Markus Ball, Mark Matzas, Valesca Boisguerin, Frank Maixner, Petra Leidinger, Christina Backes, Rabab Khairat, Michael Forster, Bjorn Stade, Andre Franke, Jens Mayer, Jessica Spangler, Stephen McLaughlin, Minita Shah, Clarence Lee, Timothy T. Harkins, Alexander Sartori, Andres Moreno-Estrada, Brenna Henn, Martin Sikora, Ornella Semino, Jacques Chiaroni, Siiri Rootsi, Natalie M. Myres, Vicente M. Cabrera, Peter A. Underhill, Carlos D. Bustamante, Eduard Egarter Vigl, Marco Samadelli, Giovanna Cipollini, Jan Haas, Hugo Katus, Brian D. O'Connor, Marc R. J. Carlson, Benjamin Meder, Nikolaus Blin, Eckart Meese, Carsten M. Pusch, and Albert Zink, 'New Insights into the Tyrolean Iceman's Origin and Phenotype as Inferred by Whole-Genome Sequencing', *Nat Commun*, 3 (2012), 698.

20 Federico Sánchez-Quinto, Hannes Schroeder, Oscar Ramirez, María C Ávila-Arcos, Marc Pybus, Iñigo Olalde, Amhed M V. Velazquez, María Encina Prada Marcos, Julio Manuel Vidal Encinas, Jaume Bertranpetit, Ludovic Orlando, M. Thomas P Gilbert, and Carles Lalueza-Fox, 'Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers', *Current biology : CB*, 22 (2012), 1494-99.

21 John Novembre, and Matthew Stephens, 'Interpreting Principal Component Analyses of Spatial Population Genetic Variation', *Nat Genet*, 40 (2008), 646-49.

22 David Reich, Kumarasamy Thangaraj, Nick Patterson, Alkes L. Price, and Lalji Singh, 'Reconstructing Indian Population History', *Nature*, 461 (2009), 489-94.

23 D. Reich, N. Patterson, D. Campbell, A. Tandon, S. Mazieres, N. Ray, M. V. Parra, W. Rojas, C. Duque, N. Mesa, L. F. Garcia, O. Triana, S. Blair, A. Maestre, J. C. Dib, C. M. Bravi, G. Bailliet, D. Corach, T. Hunemeier, M. C. Bortolini, F. M. Salzano, M. L. Petzl-Erler, V. Acuna-Alonzo, C. Aguilar-Salinas, S. Canizales-Quinteros, T. Tusie-Luna, L. Riba, M. Rodriguez-Cruz, M. Lopez-Alarcon, R. Coral-Vazquez, T. Canto-Cetina, I. Silva-Zolezzi, J. C. Fernandez-Lopez, A. V. Contreras, G. Jimenez-Sanchez, M. J. Gomez-Vazquez, J. Molina, A. Carracedo, A. Salas, C. Gallo, G. Poletti, D. B. Witonsky, G. Alkorta-Aranburu, R. I. Sukernik, L. Osipova, S. A.

bioRxiv preprint doi: https://doi.org/10.1101/001552; this version posted December 23, 2013. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fedorova, R. Vasquez, M. Villena, C. Moreau, R. Barrantes, D. Pauls, L. Excoffier, G. Bedoya, F. Rothhammer, J. M. Dugoujon, G. Larrouy, W. Klitz, D. Labuda, J. Kidd, K. Kidd, A. Di Rienzo, N. B. Freimer, A. L. Price, and A. Ruiz-Linares, 'Reconstructing Native American Population History', *Nature*, 488 (2012), 370-4.

24 Kumarasamy Thangaraj, Gyaneshwer Chaubey, Toomas Kivisild, Alla G. Reddy, Vijay Kumar Singh, Avinash A. Rasalkar, and Lalji Singh, 'Reconstructing the Origin of Andaman Islanders', *Science*, 308 (2005), 996-96.

25 Ofer Bar-Yosef, *The Chronology of the Middle Paleolithic of the Levant*. ed. by K. Aoki T. Akazawa, and O. Bar-Yosef, *Neandertals and Modern Humans in Western Asia* (New York: Plenum Press, 1998).

26 Simon J. Armitage, Sabah A. Jasim, Anthony E. Marks, Adrian G. Parker, Vitaly I. Usik, and Hans-Peter Uerpmann, 'The Southern Route "out of Africa": Evidence for an Early Expansion of Modern Humans into Arabia', *Science*, 331 (2011), 453-56.

27 J. I. Rose, V. I. Usik, A. E. Marks, Y. H. Hilbert, C. S. Galletti, A. Parton, J. M. Geiling, V. Cerny, M. W. Morley, and R. G. Roberts, 'The Nubian Complex of Dhofar, Oman: An African Middle Stone Age Industry in Southern Arabia', *PLoS One*, 6 (2011), e28239.

28 Marc Haber, Dominique Gauguier, Sonia Youhanna, Nick Patterson, Priya Moorjani, Laura R. Botigué, Daniel E. Platt, Elizabeth Matisoo-Smith, David F. Soria-Hernanz, R. Spencer Wells, Jaume Bertranpetit, Chris Tyler-Smith, David Comas, and Pierre A. Zalloua, 'Genome-Wide Diversity in the Levant Reveals Recent Structuring by Culture', *PLoS Genet*, 9 (2013), e1003316.

29 Peter Ralph, and Graham Coop, 'The Geography of Recent Genetic Ancestry across Europe', *PLoS Biol*, 11 (2013), e1001555.

30 Siiri Rootsi, Lev A. Zhivotovsky, Marian Baldovic, Manfred Kayser, Ildus A. Kutuev, Rita Khusainova, Marina A. Bermisheva, Marina Gubina, Sardana A. Fedorova, Anne-Mai Ilumae, Elza K. Khusnutdinova, Mikhail I. Voevoda, Ludmila P. Osipova, Mark Stoneking, Alice A. Lin, Vladimir Ferak, Juri Parik, Toomas Kivisild, Peter A. Underhill, and Richard Villems, 'A Counter-Clockwise Northern Route of the Y-Chromosome Haplogroup N from Southeast Asia Towards Europe', *Eur J Hum Genet*, 15 (2006), 204-11.

31 H. Shi, X. Qi, H. Zhong, Y. Peng, X. Zhang, R. Z. Ma, and B. Su, 'Genetic Evidence of an East Asian Origin and Paleolithic Northward Migration of Y-Chromosome Haplogroup N', *PLoS One*, 8 (2013), e66102.

32 O. Balanovsky, S. Rootsi, A. Pshenichnov, T. Kivisild, M. Churnosov, I. Evseeva, E. Pocheshkhova, M. Boldyreva, N. Yankovsky, E. Balanovska, and R. Villems, 'Two Sources of the Russian Patrilineal Heritage in Their Eurasian Context', *Am J Hum Genet*, 82 (2008), 236-50.

33 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.

34 Rem I. Sukernik, Natalia V. Volodko, Ilya O. Mazunin, Nikolai P. Eltsov, Stanislav V. Dryomov, and Elena B. Starikovskaya, 'Mitochondrial Genome Diversity in the Tubalar, Even, and Ulchi: Contribution to Prehistory of Native Siberians and Their Affinities to Native Americans', *American Journal of Physical Anthropology*, 148 (2012), 123-38.

35 Guido Brandt, Wolfgang Haak, Christina J. Adler, Christina Roth, Anna Szécsényi-Nagy, Sarah Karimnia, Sabine Möller-Rieker, Harald Meller, Robert Ganslmeier, Susanne Friederich, Veit Dresely, Nicole Nicklisch, Joseph K. Pickrell, Frank Sirocko, David Reich, Alan Cooper, Kurt W. Alt, and Consortium The Genographic, 'Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity', *Science*, 342 (2013), 257-61.