Early Neolithic genomes from the eastern Fertile Crescent

Farnaz Broushaki,1 Mark G Thomas,2 Vivian Link,3,4 Saioa López,2 Lucy van Dorp,2 Karola Kirsanow,1 Zuzana Hofmanová,1 Yoan Diekmann,2 Lara M. Cassidy,3 David Díez-del-Molino,2,6 Athanasios Kousathanas,5,4,7 Christian Sell,1 Harry K. Robson,4 Rui Martiniano,5 Jens Blöcher,1 Amelie Scheu,1,5 Susanne Kreutzer,1 Ruth Bollongino,1 Dean Bobo,7 Hossein Davudi,10 Olivia Munoz,11 Mathias Curtat,12 Kamyar Abdi,13 Fereidoun Biglari,14 Oliver E. Craig,8 Daniel G Bradley,5 Stephen Shennan,15 Krishna R Veeramah,9 Marjan Mashkour,16 Daniel Wegmann,3,4,† Garrett Hennenthal,2,4,† Joachim Burger1,4,†

1Palaeogenetics Group, Johannes Gutenberg University Mainz, 55099 Mainz, Germany. 2Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK. 3Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland. 4Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. 5Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland. 6Department of Bioinformatics and Genetics, Swedish Museum of Natural History, SE-10405, Stockholm, Sweden. 7Unit of Human Evolutionary Genetics, Institut Pasteur, 75015 Paris, France. 8BioArCh, Department of Archaeology, University of York, York YO10 5YW, UK. 9Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA. 10Department of Archaeology, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran. 11UMR 7041 ArScAn-VEPMO, Maison de l’Archéologie et de l’Ethnologie, 21 allée de l’Université, 92023 Nanterre, France. 12Department of Genetics and Evolution-Anthropology Unit, University of Geneva, 1211 Geneva, Switzerland. 13Samuel Jordan Center for Persian Studies and Culture, University of California–Irvine, Irvine, CA 92697-3370, USA. 14Paleolithic Department, National Museum of Iran, 113617111, Tehran, Iran. 15Institute of Archaeology, University College London, London WC1H 0PY, UK. 16CNRS/MNHN/SUS–UMR 7209, Archéozoologie et Archéobotanique, Sociétés, Pratiques et Environnements, Département Écologie et Gestion de la Biodiversité, 55 rue Buffon, 75005 Paris, France.

†Corresponding author. E-mail: jburger@uni-mainz.de (J.B.); g.hellenthal@ucl.ac.uk (G.H.); daniel.wegmann@unifr.ch (D.W.)

We sequenced Early Neolithic genomes from the Zagros region of Iran (eastern Fertile Crescent), where some of the earliest evidence for farming is found, and identify a previously uncharacterized population that is neither ancestral to the first European farmers nor has contributed significantly to the ancestry of modern Europeans. These people are estimated to have separated from Early Neolithic farmers in Anatolia some 46-77,000 years ago and show affinities to modern day Pakistani and Afghan populations, but particularly to Iranian Zoroastrians. We conclude that multiple, genetically differentiated hunter-gatherer populations adopted farming in SW-Asia, that components of pre-Neolithic population structure were preserved as farming spread into neighboring regions, and that the Zagros region was the cradle of eastward expansion.

The earliest evidence for cultivation and stock-keeping is found in the Neolithic core zone of the Fertile Crescent (I, 2); a region stretching north from the southern Levant through E-Anatolia and N-Mesopotamia then east into the Zagros Mountains on the border of modern-day Iran and Iraq (Fig. 1). From there farming spread into surrounding regions, including Anatolia and later Europe, southern Asia, and parts of Arabia and N-Africa. Whether the transition to agriculture was a homogeneous process across the core zone, or a mosaic of localized domestications is unknown. Likewise, the extent to which core zone farming populations were genetically homogeneous, or exhibited structure that may have been preserved as agriculture spread into surrounding regions, is undetermined.

Ancient DNA (aDNA) studies indicate that early Aegean farmers dating to c. 6,500-6,000 BCE are the main ancestors of early European farmers (3, 4), although it is not known if they were predominantly descended from core zone farming populations. We sequenced four Early Neolithic (EN) genomes from Zagros, Iran, including one to 10x mean coverage from a well-preserved male sample from the central Zagros site of Wezmeh Cave (WC1, 7,455-7,082 cal BCE). The three other individuals were from Tepe Abdul Hosein and were less well-preserved (genome coverage between 0.6 and 1.2 x) but are around 10,000 years old, and therefore are among the earliest Neolithic human remains in the world (tables S1 and S3).

Despite a lack of a clear Neolithic context, the radiocarbon inferred chronological age and palaeodietary data support WC1 being an early farmer (tables S1-S3 and fig. S7). WC1 bone collagen δ13C and δ15N values are indistinguishable from that of a securely assigned Neolithic individual from Abdul Hosein and consistent with a diet rich in cultivated C3 cereals rather than animal protein. Specifically, collagen from WC1 and Abdul Hosein is δ13C depleted compared to those from contemporaneous wild and domestic fauna from this region (5), which consumed C4 plants. Crucially, WC1 and the Abdul Hosein farmers exhibit very similar genomic signatures.

The four EN Zagros genomes form a distinct cluster in the first two dimensions of a principal components analysis (PCA; Fig. 2); they plot closest to modern-day Pakistani and...
Afghans and are well-separated from European hunter-gatherers (HG) and other Neolithic farmers. In an outgroup f3-test (6, 7) (figs. S17-S20) all four Neolithic Iranian individuals are genetically more similar to each other than to any other prehistoric genome except a Chalcolithic genome from NW-Anatolia (see below). Despite 14C dates spanning around 1,200 years, these data are consistent with all four genomes being sampled from a single eastern Fertile Crescent EN population.

Examination of runs of homozygosity (ROH) above 500 kb in length in WC1 demonstrated that he shared a similar ROH distribution with European and Aegean Neolithics, as well as modern day Europeans (Fig. 3A, B). However, of all ancient samples considered, WC1 displays the lowest total length of short ROH, suggesting he was descended from a relatively large HG population. In contrast, the ROH distributions of the HG Kotias from Georgia, and Loschbour from Luxembourg indicate prolonged periods of small ancestral population size (8).

We also developed a method to estimate heterozygosity (\(\hat{\theta}\)) in 1Mb windows that takes into account post-mortem damage and is unbiased even at low coverage (9) (Fig. 3C, D). The mean \(\hat{\theta}\) in WC1 was higher than in HG individuals (Bichon and Kotias), similar to Bronze Age individuals from Hungary and modern Europeans, and lower than ancient (10) and modern Africans. Multidimensional scaling on a matrix of centered Spearman correlations of local \(\hat{\theta}\) across the whole genome again puts WC1 closer to modern populations than to ancient foragers, indicating that both the mean and distribution of diversity over the genome is more similar to modern populations (Fig. 3E). However, WC1 does have an excess of long ROH segments (>1.6 Mb), relative to Aegean and European Neolithics (Fig. 3B). This includes several very long (7-16 Mb) ROH segments (Fig. 3A), confirmed by low \(\hat{\theta}\) estimates in those regions (Fig. 3C). These regions do not show reduced coverage in WC1 nor a reduction in diversity in other samples, with the exception of the longest such segment where we find reduced diversity in modern and HG individuals, although less extended than in WC1 (7) (Fig. 3B). This observed excess of long segments of reduced heterozygosity could be the result of cultural practices such as consanguinity and endogamy, or demographic constraints such as a recent or ongoing bottleneck (11).

The extent of population genetic structure in Neolithic SW-Asia has important implications for the origins of farming. High levels of structuring would be expected under a scenario of localized independent domestication processes by distinct populations, whereas low structure would be more consistent with a single population origin of farming or a diffuse homogeneous domestication process, perhaps involving high rates of gene flow across the entire Neolithic core zone. The ancient Zagros individuals show stronger affinities to Caucasus HGs (table S17.1) whereas Neolithic Aegeans showed closer affinities to other European HGs (tables S17.2 and S17.3). Formal tests of admixture of the form \(f3(\text{Neo}_1, \text{HG}; \text{Anatolia}_\text{Neolithic})\) were all positive with Z-scores above 15.78 (table S17.6), indicating that Neolithic NW-Anatolians did not descend from a population formed by the mixing of Zagros Neolithics and known HG groups. These results suggest that Neolithic populations from NW-Anatolia and the Zagros descended from distinct ancestral populations. Furthermore, while the Caucasus HGs are genetically closest to EN Zagros individuals, they also share unique ancestry with eastern, western, and Scandinavian European HGs (table S16.1), indicating that they are not the direct ancestors of Zagros Neolithics.

The significant differences between ancient Iranians, Anatolian/European farmers and European HGs suggest a pre-Neolithic separation. Assuming a mutation rate of \(5 \times 10^{-10}\) per site per year (12) the inferred mean split time for Anatolian/European farmers (as represented by Bar8, 4) and European hunter-gatherers (Loschbour) ranged from 33-39 kya (combined 95% CI 15-61 kya), while the preceding divergence of the ancestors of Neolithic Iranians (WC1) occurred 46-77 kya (combined 95% CI 38-104 kya) (13) (fig. S48 and tables S34 and S35). Furthermore, the European hunter-gatherers were inferred to have an effective population size (Ne) that was ~10-20% of either Neolithic farming group, consistent with the ROH and \(\hat{\theta}\) analyses.

Levels of inferred Neanderthal ancestry in WC1 are low (fig. S22 and table S21), but fall within the general trend described recently in Fu et al. (14). Fu et al. (14) also inferred a basal Eurasian ancestry component in the Caucasus HG sample Satsurblia when examined within the context of a “base model” for various ancient Eurasian genomes dated from ~45,000-7,000 years ago. We examined this base model using ADMIXTUREGRAPH (6) and inferred almost twice as much basal Eurasian ancestry for WC1 as for Satsurblia (62% versus 32%) (fig. S52), with the remaining derived from a population most similar to Ancient north Eurasians such as Mal’ta1 (15). Thus Neolithic Iranians appear to derive predominantly from the earliest known Eurasian population branching event (7).

‘Chromosome painting’ and an analysis of recent haplotype sharing using a Bayesian mixture model (7) revealed that, when compared to 170-230 modern groups, WC1 shared a high proportion (>95%) of recent ancestry with individuals from the Middle East, Caucasus and India. We also compared WC1’s haplotype sharing profile to that of three high coverage Neolithic genomes from NW-Anatolia (Bar8; Barcın, Fig. 4), Germany (LBK; Stuttgart) and Hungary (NE1; Polgár-Ferenci-hát). Unlike WC1, these Anatolian and European Neolithics shared ~60-100% of recent ances-
try with modern groups sampled from South Europe (figs. S24, S30, S32-S37, table S22).

We also examined recent haplotype sharing between each modern group and ancient Neolithic genomes from Iran (WC1) and Europe (LBK, NE1), HG genomes sampled from Luxembourg (Loschbour) and the Caucasus (KKI; Kotias), a 4.5k-year old genome from Ethiopia (Mota) and Ust'-Ishim, a 45k-year old genome from Siberia. Modern groups from S-, C- and NW-Europe shared haplotypes predominantly with European Neolithic samples LBK and NE1, and European HGs, while modern Near and Middle Eastern, as well as S-Asian samples had higher sharing with WC1 (figs. S28-S29). Modern Pakistani, Iranian, Armenian, Tajikistani, Uzbekistani and Yemeni samples were inferred to share >10% of haplotypes with WC1. This was true even when modern groups from neighboring geographic regions were added as potential ancestry surrogates (figs. S26-S27 and table S23). Iranian Zoroastrians had the highest inferred sharing with WC1 out of all modern groups (table S23). Consistent with this, outgroup f3 statistics indicate that Iranian Zoroastrians are the most genetically similar to all four Neolithic Iranians, followed by other modern Iranians (Fars), Balochi (SE-Iran, Pakistan and Afghanistan), Brahui (Pakistan and Afghanistan), Kalash (Pakistan) and Georgians (figs. S12-S15). Interestingly, WC1 most likely had brown eyes, relatively dark skin, and black hair, although Neolithic Iranians carried reduced pigmentation-associated alleles in several genes and derived alleles at 7 of the 12 loci showing the strongest signatures of selection in ancient Eurasians (3) (tables S29-S33). While there is a strong Neolithic component in these modern S-Asian populations, simulation of allele sharing rejected full population continuity under plausible ancestral population sizes, indicating some population turnover in Iran since the Neolithic (7).

Interestingly, while Early Neolithic samples from eastern and western SW-Asia differ conspicuously, comparisons to genomes from Chalcolithic Anatolia and Iron Age Iran indicate a degree of subsequent homogenization. Kuntepe6, a ~6,750 year old genome from NW-Anatolia (16), was more similar to Neolithic Iranians than any other non-Iranian ancient genome (figs. S17-S20; table S18.1). Furthermore, our male Iron Age genome (F38; 971-832 BCE; sequenced to 1.9x) from Tepe Hasanlu in NW-Iran shares greatest similarity with Kuntepe6 (fig. S21) even when compared to Neolithic Iranians (table S20). We inferred additional non-Iranian or non-Anatolian ancestry in F38 from sources such as European Neolithics and even post-Neolithic Steppe populations (table S20). Consistent with this, F38 carried a N1a sub-clade mtDNA, which is common in early European and NW-Anatolian farmers (3). In contrast, his Y-chromosome belongs to sub-haplogroup R1b1a2a2, also found in five Yamnaya individuals (17) and in two individuals from the Poltavka culture (3). These patterns indicate that post-Neolithic homogenization in SW-Asia involved substantial bidirectional gene flow between the East and West of the region, as well as possible gene flow from the Steppe.

Migration of people associated with the Yamnaya culture has been implicated in the spread of Indo-European languages (17, 18) and some level of Near Eastern ancestry was previously inferred in southern Russian pre-Yamnaya populations (3). However, our analyses suggest that Neolithic Iranians were unlikely to be the main source of Near Eastern ancestry in the Steppe population (table S20), and that this ancestry in pre-Yamnaya populations originated primarily in the west of SW-Asia.

We also inferred shared ancestry between Steppe and Hasanlu Iron Age genomes that was distinct from EN Iranians (table S20) (7). In addition, modern Middle Easterners and South Asians appear to possess mixed ancestry from ancient Iranian and Steppe populations (tables S19 and S20). However, Steppe-related ancestry may also have been acquired indirectly from other sources (7) and it is not clear if this is sufficient to explain the spread of Indo-European languages from a hypothesized Steppe homeland to the region where Indo-Iranian languages are spoken today. On the other hand, the affinities of Zagros Neolithic individuals to modern populations of Pakistan, Afghanistan, Iran, and India is consistent with a spread of Indo-Iranian languages, or of Dravidian languages (which includes Brahui), from the Zagros into southern Asia, in association with farming (19).

The Neolithic transition in SW-Asia involved the appearance of different domestic species, particularly crops, in different parts of the Neolithic core zone, with no single center (20). Early evidence of plant cultivation and goat management between the 10th and the 8th millennium BCE highlight the Zagros as a key region in the Neolithisation process (7). Given the evidence of domestic species movement from East to West across SW-Asia (21), it is surprising that EN human genomes from the Zagros are not closely related to those from NW-Anatolia and Europe. Instead they represent a previously undescribed Neolithic population. Our data show that the chain of Neolithic migration into Europe does not reach back to the eastern Fertile Crescent, also raising questions about whether intermediate populations in southeastern and Central Anatolia form part of this expansion. On the other hand, it seems probable that the Zagros region was the source of an eastern expansion of the SW-Asian domestic plant and animal economy. Our inferred persistence of ancient Zagros genetic components in modern day S-Asians lends weight to a strong demic component to this expansion.

REFERENCES AND NOTES
1. S. Riehl, M. Zeidi, N. J. Conard, Emergence of agriculture in the foothills of the Zagros Mountains of Iran, Science 341, 65–67

First release: 14 July 2016  www.sciencemag.org  (Page numbers not final at time of first release) 3


François de Recherche en Iran (October 2015). FBI, MM, OM and HD thank the National Museum of Iran and especially Dr. Jebrael Nokandeh, director of National Museum of Iran. We thank Nick Patterson for early access to the latest version of qpgraph. Accession numbers: Mitochondrial genome sequences are deposited in GenBank (KX353757-KX353761). Genomic data are available at ENA with the accession number PRJEB14180 in BAM format. Iranian Zoroastrian and Fars genotype data are available in plink format at figshare:

https://dx.doi.org/10.6084/m9.figshare.3470135

SUPPLEMENTARY MATERIALS
www.sciencemag.org/cgi/content/full/science.aaf7943/DC1
Materials and Methods
Figs. S1 to S52
Tables S1 to S37
References (22–162)
Animation S1

7 April 2016; accepted 5 July 2016
Published online 14 July 2016
10.1126/science.aaf7943
Fig. 1. Map of prehistoric Neolithic and Iron Age Zagros genome locations. Colors indicate isochrones with numbers giving approximate arrival times of the Neolithic culture in years BCE.
Fig. 2. PCA plot of Zagros, European, and Near and Middle Eastern ancient genomes. Comparing ancient and modern genomes, Neolithic Zagros genomes form a distinct genetic cluster close to modern Pakistani and Afghan genomes but distinct from other Neolithic farmers and European hunter-gatherers. See Animation S1 for an interactive 3D version of the PCA including the third principal component.
Fig. 3. Level and structure of ancient genomic diversity. (A) Total length of the genome in different ROH classes; shades indicate the range observed among modern samples from different populations and lines indicate the distributions for ancient samples. (B) The total length of short (<1.6Mb) vs long (≥1.6Mb) ROH. (C) Distribution of heterozygosity (θ) inferred in 1Mb windows along a portion of chromosome 3 showing the longest ROH segment in WC1. Solid lines represent the MLE estimate, shades indicate the 95% confidence intervals and dashed lines the genome-wide median for each sample. (D) Distribution of heterozygosity (θ) estimated in 1Mb windows across the autosomes for modern and ancient samples. (E) Similarity in the pattern of heterozygosity (θ) along the genome as obtained by a PCA on centered Spearman correlations. Ancient - Bich: Bichon, Upper Palaeolithic forager from Switzerland; KK1: Kotias, Mesolithic forager from Georgia; WC1: Wezmeh Cave, Early Neolithic farmer from Zagros; Mota: 4,500 year old individual from Ethiopia; BR2: Ludas-Varjú-dúló, Late Bronze Age individual from Hungary. Modern - YRI: Yoruban, W-Africa; TSI: Tuscans, Italy; PJL: Punjabi, Pakistan; GBR: British.
Fig. 4. Modern-day peoples with affinity to WC1. Modern groups with an increasingly higher (respectively lower) inferred proportion of haplotype sharing with the Iranian Neolithic Wezmeh Cave (WC1, 7,455–7,082 cal BCE, blue triangle) compared to the Anatolian Neolithic Barcin genome (Bar8: 6,212–6,030 cal BCE, red triangle) are depicted with an increasingly stronger blue color (respectively red color). Circle sizes illustrate the relative absolute proportion of this difference between WC1 versus Bar8. The key for the modern group labels is provided in table S24.
Early Neolithic genomes from the eastern Fertile Crescent


published online July 14, 2016